The American Economic Review

ARTICLES

391
.001
77

MARCH 1980

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

Officers

President
MOSES ABRAMOVITZ
Stanford University

President-Elect
WILLIAM J. BAUMOL
Princeton University and New York University

Vice Presidents
CARL F. CHRIST
Johns Hopkins University
H. GREGG LEWIS
Duke University

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review
GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature
MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee
ROBERT W. CLOWER
University of California-Los Angeles
MARINA v. N. WHITMAN
General Motors Corporation
HENRY J. AARON
Department of Health, Education, and Welfare
ZVI GRILICHES
Harvard University
MARTIN S. FELDSTEIN
National Bureau of Economic Research, Inc. and Harvard University
ROBERT E. LUCAS, JR.
University of Chicago

Ex Officio Members
TJALLING KOOPMANS
Yale University
ROBERT M. SOLOW
Massachusetts Institute of Technology

• Published at George Banta Co., Inc. Menasha, Wisconsin. The publication number is ISSN 0002-8282.

• *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Effective April 1, 1980, Association membership dues for 1978, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$28.25 for regular members with rank of assistant professor or lower, or with annual income of $13,800 or less;

$34.50 for regular members with rank of associate professor, or with annual income of $13,800 to $23,000;

$40.25 for regular members with rank of full professor, or with annual income above $23,000;

$14.25 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $43.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $4.75 to cover extra postage.

• Correspondence relating to the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

• Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212.

P17221

Vol. 70
No. 1-2

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

ALBERT ANDO
ELIZABETH E. BAILEY
PRANAB K. BARDHAN
DAVID E. BRADFORD
PETER A. DIAMOND
W. ERWIN DIEWERT
RUDIGER DORNBUSCH
MARTIN S. FELDSTEIN
WILLIAM H. OAKLAND
MICHAEL PARKIN
ROY RADNER
RICHARD W. ROLL
F. M. SCHERER
NANCY L. SCHWARTZ
A. MICHAEL SPENCE
JEROME STEIN
WILLIAM S. VICKREY
S. Y. WU

• Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to George H. Borts, Managing Editor, Box Q, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $15 for members; $30 for non-members. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1980. All rights reserved.

March 1980

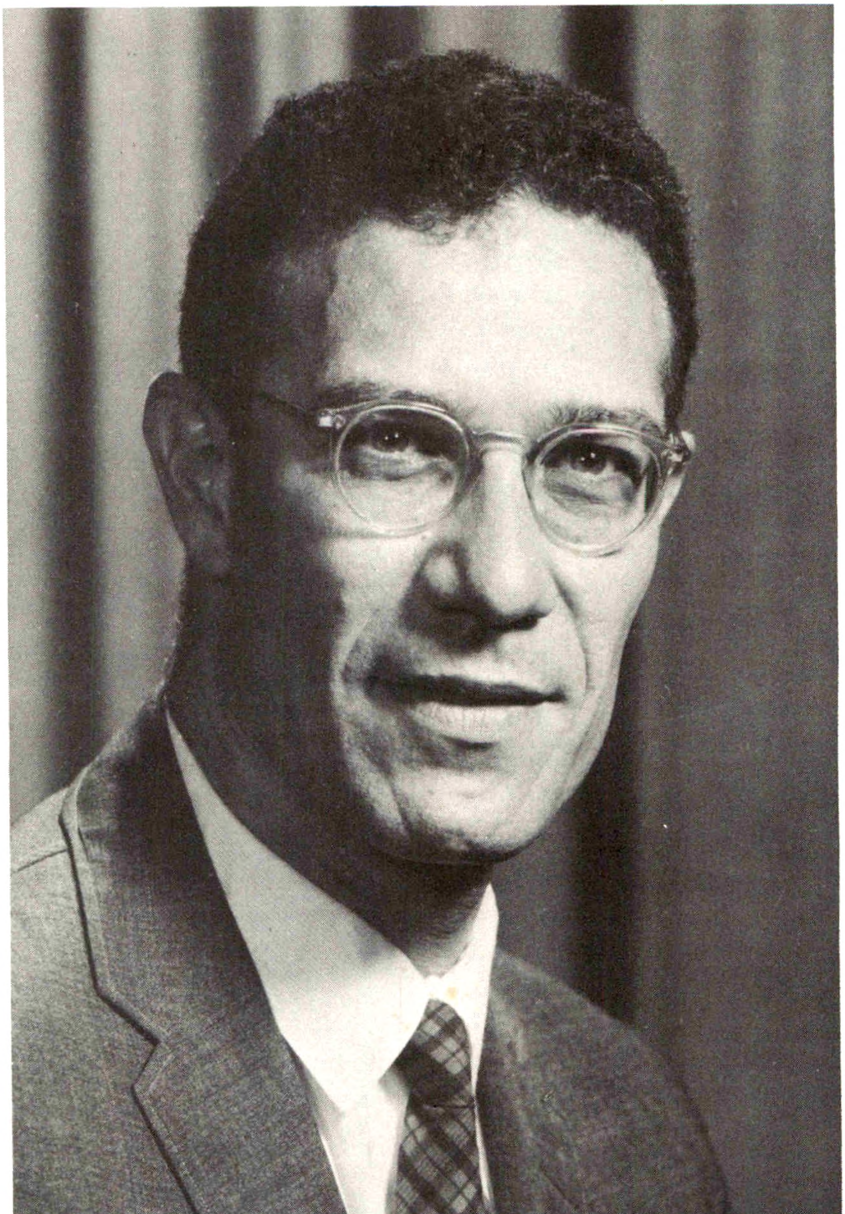
VOLUME 70, NUMBER 1

391
.001

Articles

P17321

Shorter Papers



Robert M. Solow

On Theories of Unemployment

By ROBERT M. SOLOW*

There is a long-standing tension in economics between belief in the advantages of the market mechanism and awareness of its imperfections. Ever since Adam Smith, economists have been distinguished from lesser mortals by their understanding of and —I think one has to say—their admiration for the efficiency, anonymity, and subtlety of decentralized competitive markets as an instrument for the allocation of resources and the imputation of incomes. I think we all know this; for confirmation one can look at the results of a paper (James Kearl et al.) presented at the last annual meeting, reporting the responses of professional economists to a sort of survey of technical opinion. The propositions which generated the greatest degree of consensus were those asserting the advantages of free trade and flexible exchange rates, favoring cash transfers over those in kind, and noting the disadvantages of rent controls, interest rate ceilings, and minimum wage laws.

Views on these policy issues did not seem to represent mere conservative ideology: half of the respondents agreed and another 30 percent agreed "with provisions" that redistribution of income (presumably toward the poorest) is a legitimate function of government policy. The profession's reservations about rent control, interest rate ceilings, and minimum wage laws do not appear to reflect a rejection of the goals of those measures, but rather a feeling that nonprofessionals simply do not understand fully the consequences, often unexpected and undesired, of messing around with the market mechanism. Most of us are conscious of a conflict that arises in our minds and consciences because, while we think it is usually a mistake to fiddle the price system to achieve distributional goals, we realize that the public and the political process are perversely more willing to do that than to make the direct transfers we would prefer. If we oppose all distorting transfers, we end up opposing transfers altogether. Some of us seem to welcome the excuse, but most of us feel uncomfortable. I don't think there is any very good way to resolve that conflict in practice.

Simultaneously, however, there is an important current in economics that focuses on the flaws in the price system, the ways that real markets fail because they lack some of the characteristics that make idealized markets so attractive. I think that outsiders, who tend to see economists as simple-minded marketeers, would be astonished to learn how much of the history of modern economic analysis can be written in terms of the study of the sources of market failure. The catalog runs from natural and artificial monopoly, to monopolistic competition, to the importance of public goods and externalities of many other kinds, to—most recently—a variety of problems connected with the inadequate, imperfect, or asymmetric transmission of information and with the likelihood that there will simply be no markets for some of the relevant goods and services.

Even the vocabulary can be revealing. Market "imperfection" suggests a minor blemish of the sort that can make the purchase of "irregular" socks a bargain. Market "failure" sounds like something more serious. To take a more subtle example, I mentioned that one kind of flaw in the system can be the absence of certain markets. The common generic term for the reason why markets are missing is "transaction costs." That sounds rather minor, the sort of thing that might go away in due course as accounting and information

*Presidential address delivered at the ninety-second meeting of the American Economic Association, December 29, 1979, Atlanta, Georgia. Like most people, I get by with a little help from my friends, in this case especially Paul Samuelson, George Akerlof, Arnold Kling, and James Tobin.

processing get cheaper. But some of the cases of missing markets really go much deeper. The fact that distant future generations can not participate directly in the markets for nonrenewable resources will not be remedied by improvements in communication. Nor are the residents of densely populated areas ever likely to be able to dicker effectively with the dozens or hundreds of sources of barely traceable pollutants whose health effects, if any, cumulate over many years.

There is a large element of Rohrschach test in the way each of us responds to this tension. Some of us see the Smithian virtues as a needle in a haystack, as an island of measure zero in a sea of imperfections. Others see all the potential sources of market failure as so many fleas on the thick hide of an ox, requiring only an occasional flick of the tail to be brushed away. A hopeless eclectic without any strength of character, like me, has a terrible time of it. If I may invoke the names of two of my most awesome predecessors as President of this Association, I need only listen to Milton Friedman talk for a minute and my mind floods with thoughts of increasing returns to scale, oligopolistic interdependence, consumer ignorance, environmental pollution, intergenerational inequity, and on and on. There is almost no cure for it, except to listen for a minute to John Kenneth Galbraith, in which case all I can think of are the discipline of competition, the large number of substitutes for any commodity, the stupidities of regulation, the Pareto optimality of Walrasian equilibrium, the importance of decentralizing decision making to where the knowledge is, and on and on. Sometimes I think it is only my weakness of character that keeps me from making obvious errors.

The critics of the mainstream tradition are mistaken when they attribute to it a built-in Panglossian attitude toward the capitalist economy. The tradition has provided both the foundations for a belief in the efficiency of market allocations and the tools for a powerful critique. Economic analysis by itself has no way of choosing between them; and the immediate prospects for an empirically based model of a whole economy, capable of measuring our actual "distance" from the contract curve, are mighty slim. The missing link has to be a matter of judgment—the Rohrschach test I spoke of a minute ago. For every Dr. Pangloss who makes the ink blot out to be of surpassing beauty, give or take a few minor deviations—the second-best of all possible worlds, you might say—there is a Candide to whom it looks a lot like an ink blot. Maybe there are more Panglosses than Candides. But that was true in Voltaire's time too—just before the French Revolution, by the way—and has more to do with the state of society than with the nature of economics.

The tension between market efficiency and market failure is especially pointed in discussions of the working of the labor market, for obvious reasons. The labor market connects quickly with everything else in the economy and its performance matters more directly for most people than that of any other market. Moreover, the labor market's own special pathology, unemployment, is particularly visible, particularly unsettling, and particularly frustrating. The fuse leading from theory to policy in this field is short, and has been known to produce both heat and light throughout much of the history of economics.

Contemporary macro-economic theory, though apparently full of technical novelties, has revived many of the old questions in only slightly different form. One of the points I want to make is that underneath the theoretical innovations—some of which are interesting and important—the basic controversial issues that come to the surface are the same ones that occupied earlier literature. The most important among them is really the old tension between market efficiency and market failure. Should one think of the labor market as mostly clearing, or at worst in the process of quick return to market-clearing equilibrium? Or should one think of it as mostly in disequilibrium, with transactions habitually taking place at non-market-clearing wages? In that case presumably the wage structure is either not receiving any strong signals to make it

change in the right direction or is not responding to the signals it receives. My own belief in this case lies with the market-failure side. That is to say, I believe that what looks like involuntary unemployment is involuntary unemployment.

Of course that conclusion only leads to another question. If the labor market often fails to clear, we had better figure out why. There is no shortage of candidate hypotheses. Here I think it is worthwhile to insist on a commonplace: although it is natural for academic people to seek a single weighty Answer to a weighty Question, if only because it is so satisfying to find one, it is quite likely that many of the candidate hypotheses are true, each contributing a little to the explanation of labor-market failure. Now the second general point I want to make is one that I am surprised to hear myself making. While I find several of the candidate hypotheses entirely believable, I am inclined to emphasize some that might be described as noneconomic. More precisely, I suspect that the labor market is a little different from other markets, in the sense that the objectives of the participants are not always the ones we normally impute to economic agents, and some of the constraints by which they feel themselves bound are not always the conventional constraints. In other words, I think that among the reasons why market-clearing wage rates do not establish themselves easily and adjust quickly to changing conditions are some that could be described as social conventions, or principles of appropriate behavior, whose source is not entirely individualistic.

I said that I am a little surprised at myself. That is because I am generally stodgy about assumptions, and like to stay as close to the mainstream framework as the problem at hand will allow. In any case, I think that the unconventional elements in what I have to say are only part of the story. And I assure you that I am not about to peddle amateur sociology to a captive audience. All I do mean to suggest is that we may predispose ourselves to misunderstand important aspects of unemployment if we insist on modelling the buying and selling of labor within a set of background assumptions whose main merit is that they are very well adapted to models of the buying and selling of cloth. Far from advocating that we all practice sociology, I am pleasantly impressed at how much mileage you can get from the methods of conventional economic analysis if only you are willing to broaden the assumptions a little.

I

It might be interesting to have a history of the evolution of economic ideas about unemployment, and their relation both to the internal logic of the subject and to the parallel evolution of the institutions of the labor market. I am not sufficiently well read to provide that kind of survey. To make my point about the persistence of the market-efficiency market-failure tension, I took a short cut. I went back to reread Pigou's *Lapses from Full Employment*, a little book I remember having been assigned to read as a student just after the war. And that in turn sent me back to its parent book, Pigou's *Theory of Unemployment*. The Preface to *The Theory of Unemployment* is dated April 1933, after a decade of poor performance and relatively high unemployment in Great Britain, well into the Great Depression, and before the publication of the *General Theory*. The Preface to *Lapses from Full Employment* (another example of a revealing vocabulary) is dated November 1944, after five years of the war that put an end to the depression, and well after the appearance of the *General Theory*. That seemed like an interesting approach to the historical question, because current controversies in macro-economic theory are often described as a debate between "Keynesians" and others—"monetarists," "Classicals," or " equilibrium theorists" — and because Pigou, besides being a great economist, was in particular the embodiment of the Marshallian tradition, the leading figure in the "classical economics" that the Keynesian revolution was explicitly intended to overthrow.

Lapses makes interesting rereading. It emphasizes the money wage, whereas its prede-

cessor was written almost entirely in terms of the real wage. The general macro-theoretic framework, in which the discussion of the labor market is embedded, clearly has an eye on Keynes. The underlying model could be *IS-LM* without doing much violence to the argument. There are little anachronisms: Pigou tends to think of the interest rate as being determined in the goods market (by Savings = Investment) and nominal income as being determined by the demand for money. Today we take simultaneity seriously, but the *General Theory* more or less speaks as if real output is determined in the goods market and the interest rate by liquidity preference. After what is to me a confusing description of a Keynesian low-level liquidity-trap equilibrium, Pigou invokes the Pigou effect to explain why the low level might not be as low as all that and then, characteristically, remarks that none of it is very important in practice anyway. All this is relevant here only as background for the treatment of the labor market.

Pigou says the obvious thing first, and I agree that it is the first thing to say: if there is "thorough-going competition" among workers, then the only possible equilibrium position is at full employment. That is little more than a definition of equilibrium. He is aware that he is taking a lot of dynamics for granted. Expectations of falling wages could perversely reduce the demand for labor; and he discusses the possibility that under some conditions, with the interest rate at its practical floor, nominal wage rates and prices may chase each other down and thus prevent the real-wage adjustment needed for an increase in employment. (This is where the Pigou effect makes its appearance, of course.)

It is what comes next that interests me. It is obvious to Pigou, writing in 1944, that the labor market does not behave as if workers were engaged in thorough-going competition for jobs. With the common sense that seems somehow to have escaped his modern day successors, he wonders why it does not. And he discusses three or four of the institutional factors that a reasonable person would mention even now as obstacles to the classical functioning of the labor market.

First of all, he realizes that the labor market is segmented. Not everyone in it is in competition with everyone else. I am not referring here to the obvious fact that abilities, experience, and skills differ, so that unemployed laborers can not compete for the jobs held by craftsmen. That fact of life merely reminds us that "labor" is not a well-defined homogeneous factor of production. Even within skill categories or occupational groups, however, workers have ties to localities, to industries, to special job classifications, even to individual employers. These ties can be broken, but not easily. It is interesting to me that even the *Theory of Unemployment* of 1933 devotes a lot of space to the analysis of a labor market in which there are many "centers of employment"—to use the neutral term chosen by Pigou to describe segmentation of the labor market—between which mobility is absent or slow. Of course he observes that even in a completely segmented labor market, if there is thorough-going competition within segments, full employment will be the rule, although there may be wage differentials between centers of employment for otherwise identical workers. I think that the fact of segmentation is very important, not only because it limits the scope of competition but because its pervasiveness suggests—though it can not prove—that habit and custom play a large role in labor market behavior. From the prominence that he gives it, I gather that Pigou might have agreed.

A second factor, which has been more often discussed, is trade unionism. Pigou does not have very much to say about collective bargaining, but what he says makes sense.

> Of course, these agencies in their decisions have regard to the general state of the demand for labour; they will have no wish to set wage rates so high that half the people of the country are thrown out of work. Nevertheless, there is reason to believe that they do not have regard to demand conditions in such degree as would be necessary to secure, as thorough-going competition would do, the establishment of full employment. [1945, p. 26]

Later on in the book, Pigou makes an observation that is not explicitly connected with collective bargaining. He does connect it with "actual life" however, and it fits organized workers very well, and perhaps others besides:

> In periods of expansion employers might be willing to agree to substantial advances in wage rates if they were confident that, when prosperity ended, they would be able to cancel them. They know, however, that in fact this will not be easy, that elaborate processes will have to be gone through, and that their work-people will put up a strong rear-guard action.... In periods of depression wage-earners, for precisely similar reasons, hold out against wage reductions, which they might be ready to concede if it were not for the difficulty that they foresee in getting them cancelled when times improve.... A widespread desire for 'safety first' helps to make wage rates sticky.
> [1945, p. 48]

These casual remarks raise more questions than they answer about the determination of nominal wages by collective bargaining. The first excerpt can be taken as a redefinition of full employment when the labor market is not competitive; the second, however, advances an account of wage stickiness and is therefore on a different footing. It would help to explain the failure of the labor market to clear on any reasonable definition, and thus provide a connection between nominal demand and real output.

The third institutional factor mentioned by Pigou has also been the subject of much analysis, past and present: the provision of unemployment insurance. There are several channels by which the availability of unemployment compensation can add to the recorded amount of unemployment. The prolongation of search is only the most obvious. My own impression is that this is currently a significant factor. As an indication of the complexity of the issues, let me just mention here that some recent research by my colleagues Peter Diamond and Eric Maskin suggests the possibility that in some environments search activity conveys a positive externality. So the optimal search strategy for the individual might provide less than the socially optimal amount of search, and unemployment compensation could be regarded as a corrective subsidy. This is a neat twist on the theme of the counterpoint between market efficiency and market failure. In any case, it can hardly be doubted that the unemployment compensation system is an important determinant of behavior on both sides of the labor market, and complicates even the definition of full employment.

The last comment of Pigou's that I want to cite is especially intriguing because it is so unlike the sort of thing that his present day successors keep saying. Already in the 1933 *Theory of Unemployment* he wrote: "...public opinion in a modern civilized State builds up for itself a rough estimate of what constitutes a reasonable living wage. This is derived half-consciously from a knowledge of the actual standards enjoyed by more or less 'average' workers.... Public opinion then enforces its view, failing success through social pressure, by the machinery of...legislation" (p. 255). A similar remark appears in *Lapses*. Such feelings about equity and fairness are obviously relevant to the setting of statutory minimum wages, and Pigou uses them that way. I think they also come into play as a deterrent to wage cutting in a slack labor market. Unemployed workers rarely try to displace their employed counterparts by offering to work for less; and it is even more surprising, as I have had occasion to point out in the past, that employers so rarely try to elicit wage cutting on the part of their laid-off employees, even in a buyer's market for labor. Several forces can be at work, but I think Occam's razor and common observation both suggest that a code of good behavior enforced by social pressure is one of them. Wouldn't you be surprised if you learned that someone of roughly your status in the profession, but teaching in a less desirable department, had written to your department chairman offering to teach your courses for less money? The fact that nominal wage rates did fall sharply during the early stages of the depression of the 1930's, and the fact that the Chrysler Corporation has been able

to negotiate concessions from the UAW certainly show that wage rates are not completely rigid. But those very instances seem to me only to confirm the importance of social convention in less extreme circumstances. After all, people have been known to try to claw their way into a lifeboat who would never dream of cheating on a lift-line.

I think I have made the case that the most eminent representative of orthodox economics in the 1940's was fully aware of the many obstacles to "thorough-going competition" among workers, that is, of the many ways in which the labor market may "fail." In particular, one cannot under those circumstances expect the labor market always to clear. Pigou certainly drew that conclusion. He says, in the Preface to *Lapses*: "Professor Dennis Robertson...has warned me that the form of the book may suggest that I am in favour of attacking the problem of unemployment by manipulating wages rather than by manipulating demand. I wish, therefore, to say clearly that this is not so" (p. v).

Pigou clearly felt the tension between market efficiency and market failure. Nevertheless, he did not come down on the side of market failure, even after the 1930's. The very title of *Lapses from Full Employment* tells us that much. Evidently he concluded that the tendency of the capitalist economy to seek (and find) its full-employment equilibrium was strong enough so that departures from full employment could be regarded as mere episodes. Is that surprising? Well, to begin with, there is no accounting for Rohrschach tests. One person's ink blot is another person's work of art. But I think there is also something more systematic to be said.

In the *Theory of Unemployment*, Pigou gives an elaborate analysis of the short-run elasticity of demand for labor. He is very careful: he allows for the elasticity of supply of complementary raw materials; he allows for the (presumably very high) price elasticity of demand for exports; he discusses the effects of discounting future returns to labor. It is a masterly attempt to get a grip on orders of magnitude. It is all based on the presumption that the only possible starting point is the elasticity of the marginal-product-of-labor curve. Let me remind you that in the old standby, two-factor Cobb-Douglas case, the elasticity of demand for labor with respect to the real wage is the reciprocal of the share of capital. Everybody's back-of-the-envelope puts the capital share at $1/4$ and the elasticity of demand for labor at 4. This is not exactly the way Pigou proceeds, but he reaches the same conclusion: the initial estimate of the elasticity is "certain to be (numerically) much larger than -1 and may well amount to -5 or more." There follow some modifications, but the conclusion remains that in times of depression, the aggregate elasticity of demand for labor with respect to the real wage "cannot, on the least favourable assumption here suggested, be numerically less than -3 and may well be larger than -4" except perhaps in the very shortest run.

For practical purposes, one would want to know the elasticity of demand with respect to the nominal wage, taking account of the likelihood that prices will follow wages down, at least partially. (Obviously if product prices fall equiproportionally with wage rates, as Keynes thought might happen in unlucky circumstances, the real wage doesn't move at all and employment will not improve.)[1] The details of Pigou's calculations do not concern us, but his conclusion does: "...we may...not unreasonably put the elasticity of the money demand for labour in times of deep depression at not less numerically than -1.5."

If I could believe that, I too could believe that the labor market generally clears. To reduce the unemployment rate by 6 percentage points is to increase employment by about 6 percent, if we ignore for this purpose the side effects that go to make up Okun's Law. If that could be accomplished by a real-wage reduction of 2 percent, or even less, that is, by foregoing one year's normal productivity increase, than I could imagine that the labor market might easily

[1]Neither Pigou nor Keynes invoked Kaldor's notion that prices can be expected to fall faster than wages in a recession with the resulting rise in real wages providing the force for recovery from the demand side, through a distributional shift toward wage incomes which generate more spending per dollar than other incomes do.

learn to adjust smoothly to fluctuations in aggregate demand. I could even imagine that workers might accept the necessary 4 percent reduction in nominal wages, in the expectation that half of it would be offset by lower prices. The trouble is that Pigou's demand elasticities are way too high. A recent econometric study by Kim Clark and Richard Freeman, based on quarterly data for *U.S.* manufacturing. 1950–76, puts the real-wage elasticity of demand for labor at about one-half, a whole order of magnitude smaller than Pigou's guess.[2] And the Clark-Freeman work is presented as revisionist, a counterweight to other estimates that are typically *lower*, averaging out at about 0.15 according to a survey by Daniel Hamermesh. To my mind, smooth wage adjustment seems intrinsically unlikely in a world with such a small demand elasticity and institutions like those sketched earlier. Nothing I read in the newspapers suggests to me that 6 percent of nonfrictional unemployment produces a threat adequate to set off a quick 12–15 percent fall in the real wage, or a drop in nominal wage rates twice as large. Sellers facing inelastic demands usually try to discourage price cutting; why should workers be different?

The modern classical school seems curiously remote from all this. When they try to explain how the equilibrium volume of employment can fluctuate as widely as actual employment does in business cycles, their only substitute for Pigou's high elasticity of demand is a high elasticity of supply (of labor) in the face of a perceived temporary opportunity for unusual gains, which in this case reflects wages that differ from average expected (discounted) future wages. In other words, People who give the vague impression of being unemployed are actually engaged in voluntary leisure. They are taking it now, planning to substitute extra work later, because they think, rightly or wrongly, that current real wages are unusually low compared with the present value of what the labor market will offer in the future. They may be responding to changes in real wages or to changes in the real interest rate.

It is astonishing that believers have made essentially no effort to verify this central hypothesis. I know of no convincing evidence in its favor,[3] and I am not sure why it has any claim to be taken seriously. It is hardly plausible on its face. Even if the workers in question have misread the future, they are merely mistaken, not confused or mystified about their own motives. It is thus legitimate to wonder why the unemployed do not feel themselves to be engaged in voluntary intertemporal substitution, and why they queue up in such numbers when legitimate jobs of their usual kind are offered during a recession.[4]

When they face the market-clearing issue at all, Pigou's successors take a rather abstract line. They regard it as inherently incredible that unexploited opportunities for beneficial trade should be anything but ephemeral—which means merely that they ignore all those human and institutional facts of which Pigou was aware. Or else they argue that one cannot believe in the failure of markets to clear without having an acceptable theory to explain why that happens. That is a remarkable precept when you think about it. I remember reading once that it is still not understood how the giraffe manages to pump an adequate blood supply all the way up to its head; but it is hard to imagine that anyone would therefore conclude that giraffes do not have long necks. At least not anyone who had ever been to a zoo. Besides, I think perfectly acceptable

[2]The Clark-Freeman estimates are based on quarterly data for aggregate *U.S.* manufacturing. Their difference from other work appears to rest on allowing wage changes to operate with a lag different from other factor prices. According to their results the lag of employment behind wage changes is quite short; it is complete in about two quarters.

[3]Just after writing those words, I received a working paper by Robert Hall which (a) concludes that the elasticity of supply of labor required to make the intertemporal-substitution hypothesis work is actually in the ballpark suggested by other facts, but (b) rejects the whole theory on other empirical grounds. I have done some further experimentation on Hall's data (with the help of Mr. Sunil Sanghvi) with results that cast doubt on the reliability of even the first conclusion. On reflection, I stand by the words in the text.

[4]I have tried to phrase that carefully. For some direct evidence, see "Jobs and Want Ads: A Look Behind the Evidence," *Fortune*, Nov. 20, 1978.

theories can indeed by constructed, as soon as one gets away from foolishly restrictive and inappropriate assumptions.

II

That brings me to the second and last general point I had hoped to make. Suppose one chooses to accept the apparent evidence of one's senses and takes it for granted that the wage does not move flexibly to clear the labor market. By the way, my own inclination is to go further and claim that commodity prices are sticky too, at least downward. But it is the persistence of disequilibrium in the labor market that I want to emphasize. How can we account for it?

There is, as I mentioned at the beginning, a whole catalog of possible models of the labor market that will produce the right qualitative properties. Since I have surveyed this literature elsewhere, I will just list a half-dozen possibilities now, with the reminder that they are not mutually exclusive alternatives.

(1) There is Keynes's idea that case-by-case resistance to wage reductions is the only way that workers can defend traditional wage differentials in a decentralized labor market. The net result is to preserve the general wage level or its trend, but that is an unintended artifact.

(2) There is a complementary hypothesis about the behavior of employers that I have proposed myself: if employers know that aggressive wage cutting in a buyer's market may antagonize the remaining work force, hurt current productivity, and make it harder to recruit high-quality workers when the labor market tightens, they will be less inclined to push their short-run advantage.

(3) Pigou realized that widely held notions of fairness, enforced by social pressure or by legislation, might have to be part of any serious account of wage determination. George Akerlof has pursued this trail further, documented the prescription of codes of good behavior in manuals of personnel practice, and showed formally that such codes of behavior can be self-enforcing if people value their reputations in the community. Obviously there are no Emily Post manuals to consult as regards the behavior of laid-off workers, but you would certainly not be astonished to learn that self-esteem and the folkways discourage laid-off workers from undercutting the wages of their still-employed colleagues in an effort to displace them from jobs. Reservation wages presumably fall as the duration of unemployment lengthens; but my casual reading suggests that this pattern shows up more in a willingness to accept lower-paid sorts of jobs than in "thorough-going competition" for the standard job. The cost to the worker of this sort of behavior is diminished by the availability of unemployment insurance. It is worth remembering that the acceptance of lower-grade jobs is itself a form of unemployment.

(4) I need only touch on the Azariadis-Baily-Gordon implicit-contract theory, because it has been much discussed in the literature. Here wage stability is a vehicle by which less-risk-averse firms provide income insurance for more-risk-averse workers, presumably in exchange for a lower average wage.[5] It is now understood that the theory works well only when workers have some source of income other than wages, unemployment compensation for instance. This is not really a disadvantage in a world with well-developed unemployment insurance systems. In any case such implicit contracts do not themselves account for unemployment. Their effect is to reduce the average amount of unemployment below the level that would occur in a simple spot market. The theory belongs in my list because I suspect it does help to account for the habit of wage inertia and therefore the vulnerability of employment to unexpected fluctuations in aggregate demand.

(5) Wherever there is collective bargaining in our economy, the standard pattern,

[5]Unemployment generated by this mechanism is, in a sense, voluntary. Workers reveal a preference for steady wages over steady employment. But the aggregate welfare cost of the system can still be reduced by stabilization policies. This comment applies equally to the social customs described in the preceding paragraph of the text. One can ask why workers cling to such costly conventions. It is the job of sociology to answer that question. But it is the job of economics to point out that, whatever the reason, the narrowly economic cost of such conventions can be reduced by the stabilization of aggregate demand.

with few exceptions, is that wage rates are specified in the contract, and the employer chooses the amount of employment. This is not exactly simple monopoly, because the union cannot set the wage schedule unilaterally. To the extent that it can, another source of wage stickiness can be identified. Under a reasonable assumption about what the union maximizes, it turns out that the only aspect of the demand for labor that has any effect on the monopoly wage is its elasticity. So if the demand curve for labor shifts down nearly isoelastically in a recession, the contractual wage will change little or not at all, and the full effect of the fall in demand will bear on employment. The amount of unemployment compensation available plays a role here too. (There is much more to be said along these lines, and Ian McDonald of the University of Melbourne and I hope to say it on another occasion.)

(6) As a last example, I recall Pigou's observation that wage changes may be seen by the parties as hard to reverse without a struggle whose duration and outcome cannot be foreseen. The resulting uncertainty causes employers to drag their feet when demand increases temporarily and workers to reciprocate when demand falls. The result is wage stickiness in the face of fluctuating employment.

Only what Veblen called trained incapacity could prevent anyone from seeing that some or all of these mechanisms do indeed capture real aspects of the modern capitalist economy. Assessing their combined significance quantitatively would be a very difficult task, and I do not pretend to be able to do that. We are all interpreting this ink blot together. Obviously I would not be giving this particular talk if I did not think that wage stickiness is a first-order factor in a reasonable theory of unemployment.

To make my position plausible, I want to try to summarize the sort of general characteristics that the labor market should have if the particular mechanisms that I have enumerated are to be important. By the way, I have no reason to believe that my list is anything like exhaustive; you may think of others. Simply to narrow the field, I have deliberately left out of account factors relating specifically to age, sex, race, and other characteristics that normally form the basis for discussions of structural unemployment as distinct from cyclical unemployment.

The sort of labor market I have in mind is segmented. It often makes sense to think of an employer or definable group of employers as facing its own labor pool. Some members of the labor pool may be unemployed, but still belong to it. Although transportation, information, and transaction costs are possible sources of segmentation, they need not be among the most important. The buildup of firm-specific or industry-specific human capital may be more fundamental, and equally a kind of mutual knowing-what-to-expect that gives both parties in the labor market a stake, a rent, in the durability of the relationship. This point is close to the distinction between auction markets and customer markets made by Arthur Okun in a different context. The labor market, at least the "primary" labor market, is a customer market; this may be one of the important facts that differentiates the primary from the secondary labor market.

A second general characteristic is the availability of some nontrivial source of nonemployment income. The obvious one is unemployment compensation, but I imagine that fringe activity ranging from hustling to home maintenance can function in much the same way. I suppose in some societies the possibility of returning temporarily to farming is now as important as it once was here. The presence of a second earner in the family can make an obvious difference. One consequence is that it becomes easier to maintain a labor pool in the presence of fluctuating employment. In addition, as I mentioned a few moments ago, several of the specific sticky-wage mechanisms in my catalog depend for their operation on this characteristic.

Third, the stability of the labor pool makes it possible for social conventions to assume some importance. There is a difference between a long-term relationship and a one-night stand, and acceptable behavior in one context may be unacceptable in the other. Presumably most conventions are adaptive, not arbitrary, but adaptiveness

may have to be interpreted broadly, so as to include pecuniary advantage but not be limited by it. Critics who deride the notion of "economic man" have a point, but usually the wrong point. Economic man is a social, not a psychological, category. There are activities in our culture in which it is socially acceptable and expected that individual pecuniary self-interest will be the overriding decision criterion: choosing a portfolio of securities, for example.[6] There are others in which it is not: choosing a mate, for example.[7] The labor market is more complicated than either, of course, and contains elements of both. Perhaps in nineteenth-century Manchester labor was bought and sold by "thorough-going competition" but I think that is unlikely to be a good approximation to contemporary wage setting. In particular, as I have emphasized, there is nothing in the data or in common observation to make you believe that moderate excess supply will evoke aggressive wage cutting on either side of the labor market.

III

I draw two conclusions from this whole train of thought, one about economics and the other about the economy.

About economics: it need not follow that we old dogs have to learn a lot of new tricks. It still seems reasonable to presume that agents do the best they can, subject to whatever constraints they perceive. But in some contexts the traditional formulations of the objective function and constraints may be inappropriate. In the labor market, the participants are firms and groups of firms on one side, and individual workers, organized trade unions, and informally organized labor pools on the other. Grant me that all feel constrained, to some nontrivial degree, by social customs that have to do with the wage and wage-setting procedures. The result is that factor prices turn up in our equations in unfamiliar ways. Let me just mention a few examples from my earlier list of hypotheses. If Keynes was right about the conventional significance of relative wages, then ratios of wage rates appear in the objective functions on the labor side. If the current or future performance of workers depends on their feelings that wage levels are fair, then wage rates appear in the production functions constraining firms. If the individual worker's utility function depends quite conventionally on current income, then the collective objective function of a labor pool of identical workers might reasonably be a weighted average of the utility of the wage and the utility achievable when unemployed, with weights equal to the employment and unemployment fractions. This objective function contains both wage and volume of employment as arguments; and it has the interesting property that the marginal rate of substitution between wage rate and employment can depend very sensitively on the size of the unemployment insurance benefit. Constrained maximization and partial or complete reconciliation in the market can still be the bread and butter of the macro theorist. Spread with more palatable behavior assumptions, they may make a tastier sandwich, and stick to the ribs.

About the economy: if the labor market is often not in equilibrium, if wages are often sticky, if they respond to nontraditional signals, then there is a role for macro policy and a good chance that it will be effective. Equilibrium theories that conclude the opposite may conceivably turn out to have the right answer, but they simply assume what they purport to prove. It is not my argument that standard textbook policy prescriptions are bound to be right. That has to be worked out case by case. All I do claim is that a reasonable theory of economic policy ought to be based on a reasonable theory of economic life.

[6]The emotion aroused by the case of South Africa strikes me as one of those extreme exceptions that proves the rule.

[7]In Gary Becker's defense, I should point out that he does not assume cash income to be the decisive motive in courtship.

REFERENCES

G. Akerlof, "The Case Against Conservative Macroeconomics: An Inaugural Lecture,"

Economica, Aug. 1979, *46*, 219–37.

C. Azariadis, "Implicit Contracts and Unemployment Equilibria," *J. Polit. Econ.*, Dec. 1975, *83*, 1183–202.

M. N. Baily, "Wages and Employment under Uncertain Demand," *Rev. Econ. Stud.*, Jan. 1974, *41*, 37–50.

K. Clark and R. Freeman, "How Elastic is the Demand for Labor," Nat. Bur. Econ. Res. work. Paper no. 309, Cambridge, Mass., Jan. 1979.

P. Diamond and E. Maskin, "Externalities and Efficiency in a Model of Stochastic Job Matching," working paper, Mass. Inst. Technology, forthcoming.

D. F. Gordon, "A Neo-Classical Theory of Keynesian Unemployment," *Econ. Inquiry*, Dec. 1974, *12*, 431–59,

R. Hal, "Labor Supply and Aggregate Fluctuations," Nat. Bur. Econ. Res. work. paper no. 385, Stanford, Aug. 1979.

D. Hamermesh, "Econometric Studies of Labor Demand and their Applications to Policy Analysis," *J. Hum. Resources*, Fall 1976, *11*, 507–25.

J. Kearl, C. Pope, G. Whiting and L. Wimmer, "A Confusion of Economists?," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 28–37.

A. Okun, "Inflation: Its Mechanics and Welfare Costs," *Brookings Papers*, Washington 1975, *2*, 351–90.

A. C. Pigou, *The Theory of Unemployment*, London 1933.

______, *Lapses from Full Employment*, London 1945.

R. Solow, "Alternative Approaches to Macroeconomic Theory: A Partial View," *Can. J. Econ.*, Aug. 1979, *12*, 339–54.

Inflationary Expectations, Economic Activity, Taxes, and Interest Rates

By VITO TANZI*

An empirical test of Irving Fisher's hypothesis about the impact of inflationary expectations on nominal rates of interest requires the availability of a variable that measures inflationary expectations. In the past such a variable was proxied by using distributed lags on past prices. A few years ago William Gibson used a series based on directly observed price expectations (Joseph Livingston's survey data). Recognizing that observed price expectations may contain various types of errors that might make them differ from the true, and unobservable, price expectation variable, Kajal Lahiri recently combined the information from observed price expectations with that from past rates of inflation to derive new estimated price expectations variables. In doing this he used three expectation hypotheses previously tested by Stephen Turnovsky—distributed lag, adaptive, and extrapolative—and one developed by Jacob Frenkel. The equations for short-term (i.e., six-month) expectations corresponding to each of these hypotheses are, respectively:[1]

$$\Pi_t = \sum_{i=0}^{\infty} w_i P_{t-i} \tag{1}$$

$$\Pi_t = w_0 + w_1 \Pi_{t-2} + w_2 P_t \tag{2}$$

$$\Pi_t = w_0 + w_1 P_t + w_2 (P_t - P_{t-2}) \tag{3}$$

$$\Pi_t = w_0 + w_1 \Pi_t^{12} + w_2 \Pi_{t-2} + w_3 P_t \tag{4}$$

In these equations P_t is the actual known rate of inflation (on an annual basis) for a period *ending* at the time the price forecasts are made and of the same length as the forecast. The term Π_t is the expected rate of inflation (also on an annual basis) based on the surveys, for a six-month period *beginning* at time t, while Π_t^{12} is the equivalent rate but for a twelve-month period. This twelve-month rate of expectation is assumed to reflect Frenkel's longer-term expectation variable.[2]

The least squares estimation of these four equations provided Lahiri with four estimations of the inflationary expectations, Π_t^D, Π_t^A, Π_t^E, Π_t^F (i.e., distributed lag, adaptive, extrapolative, and Frenkel's), which are partly based on the forecasts in the surveys and partly based on the actual behavior of prices in the recent past. Using a two-stage least square procedure, these estimated inflationary expectations were then used to test Fisher's theory. To summarize, six tests of Fisher's theory have been suggested. These are shown below; R_t represents the nominal rate of interest.

$$R_t = a + b\Pi_t \tag{5}$$

$$R_t = a + b\Pi_t^D \tag{6}$$

$$R_t = a + b\Pi_t^A \tag{7}$$

$$R_t = a + b\Pi_t^E \tag{8}$$

$$R_t = a + b\Pi_t^F \tag{9}$$

$$R_t = a + b_1 P_t + b_2 P_{t-1} + b_3 P_{t-2} \tag{10}$$

Equation (5) follows Gibson's approach; equations (6)–(9) are Lahiri's; and equation (10) is one version of the more traditional distributed lags approach using past prices.

*Chief, Tax Policy Division, fiscal affairs department, International Monetary Fund. The opinions expressed in this paper are my own and do not represent official IMF views. I have benefited from comments received from Ke-young Chu, A. Feltenstein, J. A. Frenkel, Mohsin S. Khan, George F. Kopits, and an anonymous referee. I am grateful to all of them while remaining responsible for any remaining errors. Chris Wu provided very able statistical assistance.

[1]The rationale behind these various expectation hypotheses can be found in the cited papers by Turnovsky, Frenkel, and Lahiri.

[2]Observed expectations for periods longer than twelve months are not available.

The specific formulation shown in equation (10) follows Robert Gordon's derivation of inflationary expectations for his study of the Phillips curve. Following the basic Fisherian theory all of these equations assume that the real rate of interest does not play a role in determining the nominal rate.

The objectives of the present paper are the following: (a) to reestimate equations (5)–(9) on the belief that the data used by Gibson and Lahiri were faulty and that these errors may have affected their results; in the process of doing so the period covered will be extended beyond 1970 to 1975; (b) to present and test a theory that changes in economic activity affect the real rate of interest so that these changes must be included in the basic Fisherian equation; finally, (c) to test the hypothesis that when there are income taxes the increase in nominal rates of interest must exceed inflationary expectations. Two variables never used before for this purpose are used to test these theories.

I. Reestimation of Basic Equations

There are several reasons why it is worthwhile to reestimate the above equations. First, it is worthwhile to extend the period. More importantly, however, it is necessary to use correct data to be sure that the results are valid. In a recent paper, John Carlson carefully reworked Livingston's data to remove errors and has published the corrected series.[3] Carlson's revised series (for the *CPI*) is the one used in this paper.[4] Furthermore, Lahiri used the yield on three-month Treasury Bills in connection with six-month inflationary expectations to be able to cover the 1952–70 period. Thus, to gain degrees of freedom, he introduced another potentially important error.

In this paper I will use six-month inflationary expectations together with the yield on six-month Treasury Bills, but I will also use twelve-month expectations together with the yield on twelve-month bills.[5] For the former, the period covered will be June 1959–December 1975 while for the latter it will be June 1952–December 1975, periods for which the data are available. Table 1 provides the regression equations explaining the formation of inflationary expectations. For the six-month period the four alternative hypotheses are used while for the twelve-month period only three of the hypotheses could be tested.[6] In all cases the $\bar{R}^2$s are very high, but they are much higher for six-month expectations than for twelve months. The Durbin-Watson (*D.W.*) statistics were also much higher for the equations explaining six-month expectations. Frenkel's hypothesis gave the best result on the $\bar{R}^2$ criterion; on the other hand, it had the lowest *D.W.* value. After Frenkel's, the adaptive hypothesis, corresponding to equation (2) above, gave the best results: it had the highest *D.W.* value for both six- and twelve-month expectations and very high $\bar{R}^2$s. In conclusion, the various hypotheses did quite well in explaining the formation of observed expectations. The results in Table 1 are far superior to those obtained by Lahiri (see his Table 1, p. 127).

Table 2 shows the regression estimations of equations (5)–(10) described in the previous section. These are the basic tests of the Fisherian hypothesis and have been corrected for autocorrelation using a first-order Cochrane-Orcutt correction factor. The uncorrected equations had *D.W.* statistics ranging from 0.55 to about 1.00. The first regression equation in each group corresponds to equation (5). The last is based on Gordon's hypothesis, and thus corresponds to equation (10). For six-month bills, the middle four regression equations correspond to equations (6) to (9) and are based on a

[3]See Carlson (1977a, especially Table 1, pp. 33–34). As Carlson has convincingly shown, in deriving estimates of the expected rates of inflation from Livingston's survey data, Gibson made potentially important errors which were not correct by Lahiri since the latter just used Gibson's data.

[4]The use of the incorrect series implies also that wrong past prices were used by Lahiri in the estimation of equations (1)–(4) and (6)–(9).

[5]Lahiri's own analysis was limited to six-month Treasury Bills.

[6]The testing of Frenkel's hypothesis in connection with twelve-month expectations would have required the use of observed expectations for periods much longer than twelve months.

TABLE 1—ORDINARY LEAST SQUARES TO EXPLAIN THE FORMATION OF EXPECTATIONS

Expectation Hypothesis	Estimates	$\bar{R}^2$	D.W.	SEE
Six-Month Expectations: June 1959–December 1975				
Distributed	$\Pi_t = 0.2693 + 0.4312P_t + 0.0464P_{t-1} + 0.1619P_{t-2}$ (1.75) (7.07)[a] (0.55) (2.64)[b]	0.929	1.480	0.513
Adaptive	$\Pi_t = 0.1776 + 0.4173P_t + 0.3715\Pi_{t-2}$ (1.23) (9.81)[a] (5.12)[a]	0.942	1.875	0.466
Extrapolative	$\Pi_t = 0.2701 + 0.6396P_t - 0.1849(P_t - P_{t-2})$ (1.77) (20.06)[a] (4.21)[a]	0.931	1.619	0.507
Frenkel	$\Pi_t = 0.2230 + 1.0461\Pi_t^{12} + 0.0162P_t - 0.0408\Pi_{t-2}$ (5.38)[a] (20.56)[a] (0.72) (1.50)	0.996	1.310	0.118
Twelve-Month Expectations: June 1952–December 1975				
Distributed	$\Pi_t = -0.0148 + 0.6790P_t - 0.2647P_{t-1} + 0.2760P_{t-2}$ (0.08) (5.79)[a] (1.36) (2.04)[b]	0.833	0.722	0.820
Adaptive	$\Pi_t = 0.0953 + 0.3235P_t + 0.5429\Pi_{t-2}$ (0.66) (4.87)[a] (5.49)[a]	0.894	0.928	0.654
Extrapolative	$\Pi_t = 0.0257 + 0.6733P_t - 0.1238(P_t - P_{t-2})$ (0.14) (13.79)[a] (1.60)	0.829	0.565	0.828

Note: The numbers in parentheses are the coefficients divided by their standard errors; $\bar{R}^2$ is adjusted R^2; *D.W.* is the Durbin-Watson statistic; *SEE* is the standard error of the estimate. For twelve-month expectations Frenkel's hypothesis could not be tested due to lack of expectations extending substantially beyond twelve months.

[a]Indicates significance at the 1 percent level.

[b]Indicates significance at the 5 percent level.

TABLE 2—LEAST SQUARES ESTIMATES OF THE EFFECT OF INFLATIONARY EXPECTATIONS ON INTEREST RATES

Expectation Hypotheses	Estimates	$\bar{R}^2$	D.W.	SEE
Six-Month Bills				
Observed	$R_t = 3.0421 + 0.6548\Pi_t$ (5.08)[a] (4.30)[a]	0.353	1.657	0.753
Distributed	$R_t = 3.1645 + 0.6367\Pi_t$ (4.87)[a] (4.27)[a]	0.365	1.606	0.660
Adaptive	$R_t = 3.1624 + 0.6313\Pi_t$ (4.82)[a] (4.11)[a]	0.346	1.620	0.674
Extrapolative	$R_t = 3.1815 + 0.6286\Pi_t$ (4.92)[a] (4.34)[a]	0.373	1.608	0.654
Frenkel	$R_t = 3.2692 + 0.5932\Pi_t$ (4.82)[a] (3.64)[a]	0.290	1.633	0.709
Gordon	$R_t = 3.5594 + 0.4441P_t - 0.0876P_{t-1} - 0.0248P_{t-2}$ (6.63)[a] (5.75)[a] (1.08) (0.35)	0.491	1.968	0.660
Twelve-Month Bills				
Observed	$R_t = 2.7130 + 0.7144\Pi_t$ (7.41)[a] (6.22)[a]	0.450	1.839	0.776
Distributed	$R_t = 3.8096 + 0.2979\Pi_t$ (4.96)[a] (1.74)	0.045	1.984	0.886
Adaptive	$R_t = 3.1209 + 0.5364\Pi_t$ (5.45)[a] (3.18)[a]	0.171	1.825	0.885
Extrapolative	$R_t = 3.7014 + 0.3373\Pi_t$ (4.88)[a] (1.81)	0.049	1.945	0.888
Gordon	$R_t = 4.0078 + 0.2077P_t - 0.0830P_{t-1} + 0.0308P_{t-2}$ (4.37)[a] (1.57) (0.54) (0.23)	0.010	2.025	0.906

Notes: See Table 1.

[a]See Table 1.

two-stage least squares procedure whereby the Π_t are calculated from the equations in Table 1. These Π_t are then used as independent variables in Table 2.[7] Frenkel's hypothesis could not be tested for twelve-month expectations.

Most of the equations in Table 2 support the hypothesis of a relationship between the interest rate variable and inflationary expectations: in most cases, the *t*-values for the inflationary expectation variable are high. However, the coefficients of the expectation variables are significantly below unity indicating that not all the effect of inflationary expectations is incorporated in the nominal rate of interest or, perhaps, that the equations were misspecified. In no case are these coefficients as high as some found by Lahiri or by Gibson. For six-month bills, for example, Gibson had found for the expected inflation variable a coefficient of 0.936; for twelve-month bills he had found a coefficient of 0.91. In addition, the relationship between nominal interest rates and inflationary expectations are no longer as good as those found by Lahiri and Gibson. This coupled with high serial correlation in the original equations raises the issue of whether some important variable has not been left out of the relationship.

II. Economic Activity and Interest Rates

A. *Theory*

The assumption of a constant real rate of interest underlying the Fisherian analysis cannot be accepted once cyclical fluctuations are recognized. Yet, these fluctuations have been largely ignored by the recent literature dealing with interest rates under inflationary conditions. The impact of business cycles on the real rate of interest may come about in two ways: through their effects on inflationary expectations; or through a direct impact on the real rate that economic agents expect to pay or receive at a given moment of time. The first of these effects is important as far as the determination of *realized* real rates is concerned but is not as relevant to the Fisherian theory unless this theory is taken to explain not only expected but also realized real rates. The second, on the other hand, goes to the core of it.

Let Π_t, P_t, and R_t continue to indicate inflationary expectations, however determined, actual rate of inflation, and nominal rate of interest all for the same period t. Let $\bar{U}$ and U be the full (or natural) rate of unemployment and actual rate of unemployment, respectively; $\bar{Y}$ and Y are the full (or potential) and actual output, respectively, and r_t^e, r_t^r, r^n are the expected, realized, and long-run (or natural) real rates of interest, respectively.

From Fisher's theory,

$$R_t = r_t^e + b\Pi_t \tag{11}$$

Under the usual Fisherian assumption, Π is the only variable affecting R; furthermore, changes in Π are fully reflected in R so that the expected real rate r_t^e is implicitly assumed constant and equal to r^n. Thus the b can be dropped from this equation. Of course, for any particular period, because of random factors, the expected (or *ex ante*) real rate (equal to $R_t - \Pi_t$) can diverge from the realized (*ex post*) real rate ($R_t - P_t$). If we define $(r_t^e - r_t^r) = (P_t - \Pi_t) = D_t$, Fisher's theory requires that

$$E(D_t) = 0$$

$$E(D_t D_{t'}) = \sigma^2 \qquad \text{if } t = t'$$

$$= 0 \qquad \text{if } t \neq t'$$

On the other hand, if business fluctuations have an impact on D_t, these conditions will not be met. One hypothesis that argues for systematic changes in D_t around a long-run expected value of zero, is the one behind the so-called expectations-augmented Phillips curve (see Rudiger Dornbusch and Stanley Fischer, pp. 404–07). This hypothesis can be formulated as follows:

$$P_t = \Pi_t - \varepsilon(U - \bar{U})_t \tag{12}$$

[7]This two-stage approach is preferred to a single stage one, which would use as dependent variables the same ones as those in the equations of Table 1, mainly because it facilitates the direct testing of Fisher's hypothesis (through an analysis of the coefficients of the Π_s).

that is, the divergence between the unemployment rate and the natural rate of unemployment is likely to bring about systematic differences between actual and expected rates of inflation.

Equation (12) can be related to output rather than employment through the use of Okun's Law which can be written as

$$(13) \qquad (U-\bar{U})_t = -\alpha(Y-\bar{Y})_t \qquad \alpha>0$$

Combining equations (12) and (13) we get

$$(14) \qquad P_t = \Pi_t + \alpha\varepsilon(Y-\bar{Y})_t \qquad \alpha\varepsilon>0$$

which states that when actual output exceeds (falls short of) potential output, actual price changes will tend to be higher (lower) than expected price changes.[8]

Solving equation (14) for Π_t and combining with equation (11) we get

$$(15) \qquad R_t = r_t^e + P_t - \alpha\varepsilon(Y-\bar{Y})_t$$

This equation indicates that at full employment (i.e., $Y=\bar{Y}_t$) the nominal rate of interest will increase in line with the actual rate of inflation. In such case $r_t^e = r_t^n = r^r$ and inflation would have no effect on the real rate. However, a boom $(Y>\bar{Y})_t$ or a recession $(Y<\bar{Y})_t$ would bring about systematic effects on R_t and, of course, on the realized real rate $(R_t - P_t)$.

The above theory has emphasized the effect of business cycles on the divergence between P and Π, but has ignored possible effects on r^e. However, cyclical fluctuations can also influence the expected real rate of interest. For example, cyclical shifts in the investment function (be they due to "waves of optimism and pessimism," to "bunchings of innovations," or to other factors) cannot fail to have an impact on the credit market. In a recession, firms will liquidate inventories, postpone investment plans, repay bank loans, and so forth, so that if, as it seems likely, the demand for loanable funds falls more than the supply,[9] the credit market will turn more favorable to borrowers. In such circumstances, borrowers will expect to pay, and lenders expect to receive, lower rates of interest than in better times. The net result should be a fall in the expected real and, given inflationary expectations, in the nominal rate of interest. The opposite will be true in a boom.

We can formalize this relationship as

$$(16) \qquad r_t^e = r^n + \theta(Y-\bar{Y})_t \qquad \theta>0$$

Substituting the value of r_t^e from equation (16) into Fisher's basic equation we get

$$(17) \qquad R_t = r^n + \Pi_t + \theta(Y-\bar{Y})_t$$

This is our basic revision of the Fisherian equation and it will form the basis for the empirical tests. This equation is still expressed in terms of inflationary expectations as required by Fisher's theory but it adds the level of economic activity as a basic variable. It states that during periods of high activity $(Y>\bar{Y})_t$ nominal rates of interest will increase by more than inflationary expectations. The opposite will occur in periods of recession.

Before leaving this section it may be worthwhile to add that the full impact of fluctuations in real output on the *realized* real rate of interest can be obtained by combining equations (15) and (16). Therefore,

$$(18) \qquad R_t - P_t = r^n + (\theta - \alpha\varepsilon)(Y-\bar{Y})_t$$

Thus fluctuations in real output affect the realized real rate through two channels, namely, the impact on expectations and that on expected real rates. These effects have opposite signs so that the net effect depends on the size of the coefficients θ and $\alpha\varepsilon$. These effects are shown graphically in Figure 1.

B. *Empirical Results*

Two recent empirical studies that have tried to take into account the effect of fluctuations in real output on expected real interest rates are by J. Walter Elliott and

[8]For the theoretical justification see Dornbusch and Fisher, p. 404.

[9]The accelerator principle is likely to bring this about.

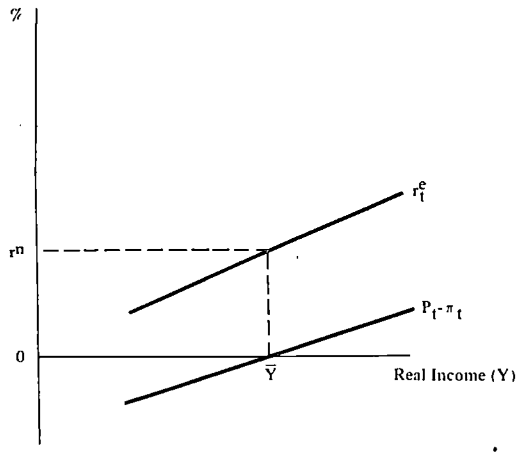


FIGURE 1

Carlson. Elliott's results were not very encouraging. Regressing the expected real rate on both the rate of real output change and real output levels Elliott found "a weak relationship with the rate of real output change and no relationship with output level"; he concluded that "a dominant business cycle influence upon the expected real rate is not apparent in the estimated series" (p. 442). Carlson's 1977b study showed that the ratio of employment to population—which he took to be a measure of economic activity—was of some help in explaining the relationship between changes in interest rates and changes in purchasing power. Eugene Fama, on the other hand, has concluded "that the largest part of the variation in nominal interest rates reflects variation in expected inflation rates..." (p. 496).

A basic issue in introducing the level of economic activity in the equations is the selection of an adequate index to reflect this activity. For our own purpose this index must be available, at least since 1952, for the same periods of the forecasts and must closely reflect what is happening to real output without a lag and without magnifying or smoothing its fluctuations. In view of the theoretical discussion of the previous section, the index that seems most appropriate for this purpose, and that has been used in this paper, is based on the difference between potential and actual *GNP*. This index, which will be called the gap G_t, is

$$G_t = \frac{\text{Actual } GNP_t - \text{Potential } GNP_t}{\text{Potential } GNP_t} \times 100$$

This index is negative during recessions and zero or positive during periods of booms. Over the period, and on a semester basis, it has ranged from a negative value of 10.44 percent to a positive one of 3.44 percent.[10]

As suggested by equation (17), the gap must be added as a new independent variable in equations (5) to (10). Before this is done, however, we must ensure that no multicollinearity exists between inflationary expectations and economic activity. The possibility of such a relationship is, of course, suggested by the Phillips curve. Consequently, the price expectation variables used in Table 1 were regressed against the economic activity variable. In most cases no significant relationship was found. In the few cases in which some relationship did exist, the $\bar{R}^2$ was very low, around 0.10 or less.

Three alternative lags were tried for the gap.[11] The first covered the same period as the price forecast; the second covered the period immediately before that of the forecast; the third was estimated for a twelve-month period centered at the time of the forecast. The first of these proved the worst. The other two gave much better results with the one centered at the time of the forecast performing a little better on the $\bar{R}^2$ criterion, while the one for the period before the forecast performed better on the *D.W.* criterion. Only the results with the lag centered at the time of the forecast have been shown.

The regression results for the new equations are shown in Table 3. They have been

[10]The potential *GNP* series used is prepared by the Council of Economic Advisers. The series for quarterly potential *GNP* used in this paper has not been published. It was kindly made available by the Council. The yearly data were published in the *Economic Report of the President*.

[11]A priori it was difficult to decide which would be the most appropriate.

TABLE 3—LEAST SQUARES ESTIMATES OF INTEREST RATE EQUATIONS WITH INDEX OF ECONOMIC ACTIVITY

Expectation Hypotheses	Estimates	$\bar{R}^2$	$D.W.$	SEE
Six-Month Bills				
Observed	$R_t = 3.1950 + 0.7822\Pi_t + 0.2057G_t$	0.610	1.833	0.667
	$(8.45)^a$ $(7.10)^a$ $(3.99)^a$			
Distributed	$R_t = 2.9892 + 0.8791\Pi_t + 0.2480G_t$	0.805	1.887	0.523
	$(11.74)^a$ $(11.14)^a$ $(5.69)^a$			
Adaptive	$R_t = 3.0305 + 0.8466\Pi_t + 0.2271G_t$	0.757	1.837	0.568
	$(10.53)^a$ $(9.68)^a$ $(4.73)^a$			
Extrapolative	$R_t = 2.9952 + 0.8724\Pi_t + 0.2435G_t$	0.807	1.901	0.526
	$(11.89)^a$ $(11.22)^a$ $(5.64)^a$			
Frenkel	$R_t = 3.1288 + 0.7899\Pi_t + 0.2042G_t$	0.615	1.809	0.629
	$(8.08)^a$ $(7.00)^a$ $(3.43)^a$			
Gordon	$R_t = 3.2829 + 0.4726P_t - 0.0032P_{t-1} + 0.049P_{t-2} + 0.2020G_t$	0.814	1.888	0.574
	$(14.87)^a$ $(6.77)^a$ (0.04) (0.73) $(4.30)^a$			
Twelve-Month Bills				
Observed	$R_t = 2.908 + 0.8268\Pi_t + 0.1969G_t$	0.618	1.999	0.698
	$(9.97)^a$ $(8.67)^a$ $(3.47)^a$			
Distributed	$R_t = 3.479 + 0.6953\Pi_t + 0.2722G_t$	0.378	2.260	0.803
	$(8.00)^a$ $(5.07)^a$ $(3.51)^a$			
Adaptive	$R_t = 3.0487 + 0.8539\Pi_t + 0.2724G_t$	0.569	1.972	0.770
	$(9.33)^a$ $(7.61)^a$ $(4.16)^a$			
Extrapolative	$R_t = 3.3092 + 0.7955\Pi_t + 0.2982G_t$	0.437	2.185	0.784
	$(8.14)^a$ $(5.76)^a$ $(3.96)^a$			
Gordon	$R_t = 3.2666 + 0.3433P_t + 0.0502P_{t-1} + 0.2034P_{t-2} + 0.3420G_t$	0.418	2.161	0.785
	$(7.68)^a$ $(3.00)^a$ (0.34) (1.71) $(4.10)^a$			

Notes: See Table 1.
[a]See Table 1.

corrected for serial correlation using a first-order Cochrane-Orcutt correction.[12] They show clearly that the introduction of the index of economic activity in the basic equations improves all the equations. First, the $\bar{R}^2$s are sharply raised. Secondly, the coefficients of the variable representing inflationary expectations have sharply increased and are now in most cases close to one. This would indicate that the Fisherian hypothesis—that a 1 percentage point rise in expected inflation will bring about a 1 percentage point increase in nominal interest rates—is essentially correct provided that there are no fluctuations in real output. When there are such fluctuations the expected real rate will change. Therefore, various nominal rates of interest can be associated with the same inflationary expectation provided that the index of economic activity varies. On the average, an increase of 1 percentage point in the gap will be associated with an increase of about one-fourth of a percentage point in the rate of interest. This aspect is relevant to the current controversy of whether cyclical fluctuations can be ignored when interest rates are used as predictors of inflation.

Table 3 shows that the level of economic activity is one of the important determinants of nominal interest rates and, by implication, of the real rates of interest that individuals expect to receive at given periods of time. If lenders have inflationary expectations equal to Π_t and yet buy Treasury Bills which pay R_t, they are expecting to receive a real rate of interest equal to the difference between R_t and Π_t. Several alternative measures for the inflationary expectation variable have been used above. By subtracting these measures from the nominal rate of interest, one can derive alternative estimates of the expected real rate of interest. These

[12]For the uncorrected equations the $D.W.$ statistics ranged between 1.00 and 1.57; otherwise, they were almost identical to the corrected ones.

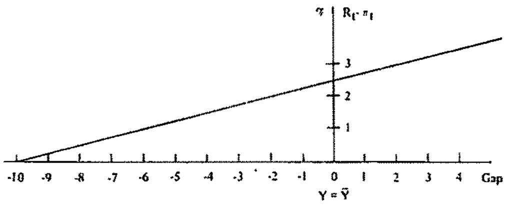


FIGURE 2

can then be directly correlated with the index of economic activity. In other words, the following equation can be estimated:

$$(19) \qquad R_t - \Pi_t = f(G_t)$$

The regression equations derived in this manner show beyond doubt the effect of economic activity on expected real rates.[13] These equations indicate that when G_t is zero the expected annual real rate of interest averages around 2.50. This expected real rate rises by about 1 percentage point when G_t is equal to 4. Thus, a negative gap of about 10 reduces the expected real rate of interest to about zero (see Figure 2).

III. Income Taxes and Interest Rates

Up to this point I have followed Fisher and his followers in ignoring the existence of income taxes and their possible impact on interest rates. However, in recent papers, several writers (see, *inter alia*, Michael Darby, Martin Feldstein, and my paper) have shown that, since interest income is taxed, a rate of inflation that resulted in an equivalent increase in the nominal rate of interest would not leave the lender in the same preinflationary real situation. Rather, the rate that would compensate the lender for inflation *and for the payment of income taxes* would have to be somewhat higher than Fisher's. For a lender who suffers from neither money nor fiscal illusion, the *ex ante* rate required to leave him in the same real preinflationary situation can be approximated by the following expression:

$$(20) \qquad R_t^* = r_t + \frac{\Pi_t}{1 - T_t}$$

where R_t^* is the lender's required rate at time t, r_t is the rate that he would have received had there been no inflationary expectation, Π_t is his expected rate of inflation, and T_t is the rate at which his interest income would be taxed.

Equation (20) is in effect equation (5) above but in a slightly different guise. It could also be written as

$$(21) \qquad R_t^* = r_t + b\Pi_t$$

However, while in equation (5) theoretical considerations (i.e., Fisher's hypothesis) had led us to expect that b should not be significantly different from one, in equation (21) theoretical considerations lead us to expect that b (which now is an estimation of $1/(1-T_t)$) ought to be greater than one. Over the 1952–75 period the average tax rate on interest income ranged from a low of 25.3 percent in 1967 to a high of 34.8 percent in 1974. The mean for the 1952–75 period was 0.32.[14] Therefore, if individuals did not suffer from money and/or fiscal illusion, the b coefficient in equation (21) should be somewhere around 1.47 (i.e., $1/(1-0.32)$).

If, over the period, the weighted average tax rate T_t had been constant and equal to the mean of 0.32, then we could test for the effect of taxes on interest rates by using the equations already estimated. In this case, we would simply test for the significance of the coefficients of the expectation variable, using the null hypothesis that they were not different from 1.47. If this test were applied, all the coefficients obtained above would fail. In all cases they would be found to be significantly lower than the expected value of 1.47. However, since the average tax rate T_t was not constant over the period, a more precise test involves adjusting, for each year,

[13]To save space these equations are not shown. They are available on request from the author.

[14]For the 1952–75 period this rate was obtained on the basis of information contained in *Statistics of Income*. This estimated series is available on request from the author.

the expectation variables by dividing them by $1-T_t$. In such a case the independent variable Π would be corrected for the effect of taxes *before* estimating the regression equations. The variable Π_t would thus be replaced by $\Pi_t/(1-T_t)$ so that the regression coefficients for this corrected variable would still be expected to be one. The application of this alternative test failed to discover any tax effect. When Π_t was replaced by $\Pi_t/(1-T_t)$ in the various regression equations, its coefficient fell significantly below the expected value of one.[15]

The inevitable conclusion is that individuals have, to a large extent, been able to see through the money veil, and thus have not suffered from money illusion; however, they have failed to see through the fiscal veil and thus have suffered from fiscal illusion. Interest rates have not adjusted enough to compensate individuals for the combined effect of inflationary expectations *and* income taxes. That individuals were unable to see through the particular fiscal illusion discussed in this section is hardly surprising when it is realized that it has been "discovered" by economists only in the past couple of years. Thus, since for most of the period covered by this analysis the relevant theory ignored the effect of income taxes, the existence of fiscal illusion does not mean that the expectations held by investors were "irrational" (see Richard Muth).

IV. Summary and Conclusions

In this paper I have reanalyzed the behavior of interest rates under inflationary conditions by using various inflationary expectation hypotheses. The results indicated that, when an inflationary expectation variable is used as the only independent variable in explaining changes in interest rates, the latter do not increase by the same amount as the former. This relationship improved substantially when a variable indicating the level of economic activity was introduced. The implication of this is that the expected real rates of interest do not remain constant but change in line with the business cycle. This result supports the view, held by Carlson and others and opposed by Fama, that interest rates alone may not be good predictors of inflation when the level of economic activity is changing rapidly.

The paper tried also to give an empirical answer to a recent and growing body of theoretical literature that maintains that income taxes must have an effect on nominal rates of interest under inflationary conditions. It was shown that during the 1952–75 period people suffered from fiscal illusion in the sense that the effect of income taxes in reducing the net-of-taxes expected real rate of interest was ignored. This result is not considered surprising since that effect is still not widely recognized even by economists. This, of course, does not mean that individuals will continue to suffer from these illusions if inflation continues.

[15]These equations are not shown in the text. They are available on request from the author.

Appendix

Π_t: observed inflationary expectations were obtained from Carlson, (1977a, Table 1, pp. 33–34). These data are based on Livingston's surveys.

R_t: Treasury Bill rates were obtained from Board of Governors of the Federal Reserve system, *Banking and Monetary Statistics, 1941–1970*, and *Annual Statistical Digest, 1971–75*.

Gap: calculated from the potential (quarterly) *GNP* series prepared by the Council of Economic Advisers.

P_t: monthly data on actual inflation rates were obtained from a computer printout from the U.S. Bureau of Labor Statistics.

T_t: the weighted-average tax rates were estimated by the author on the basis of the annual *Statistics of Income*.

REFERENCES

J. A. Carlson, (1977a) "A Study of Price Forecasts," *Annals Econ. Soc. Measure.*, June 1977, *1*, 27–52.

______, (1977b) "Short-Term Interest Rates

as Predictors of Inflation: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 469–75.

M. R. Darby, "The Financial and Tax Effects of Monetary Policy on Interest Rates," *Econ. Inquiry*, June 1975, *13*, 266–76.

Rudiger Dornbusch and Stanley Fischer, *Macroeconomics*, New York 1978, 404–07.

J. W. Elliott, "Measuring the Expected Real Rate of Interest: An Exploration of Macroeconomic Alternatives," *Amer. Econ. Rev.*, June 1977, *67*, 429–44.

E. F. Fama, "Interest Rates and Inflation: The Message in the Entrails," *Amer. Econ. Rev.*, June 1977, *67*, 487–96.

M. Feldstein, "Inflation, Income Taxes, and the Rate of Interest: A Theoretical Analysis," *Amer. Econ. Rev.*, Dec. 1976, *66*, 809–20.

J. A. Frenkel, "Inflation and the Formation of Expectations," *J. Monet. Econ.*, Oct. 1975, *1*, 403–22.

W. E. Gibson, "Interest Rates and Inflationary Expectations: New Evidence," *Amer. Econ. Rev.*, Dec. 1972, *62*, 854–65.

R. J. Gordon, "Inflation in Recession and Recovery," *Brookings Papers*, Washington 1971, *1*, 105–66.

K. Lahiri, "Inflationary Expectations: Their Formation and Interest Rate Effects," *Amer. Econ. Rev.*, Mar. 1976, *66*, 124–31.

Joseph A. Livingston, surveys published twice yearly, *Philadelphia Sunday Bulletin*, 1948–71; *Philadelphia Inquirer*, 1972 on.

J. F. Muth, "Rational Expectations and the Theory of Price Movements, *Econometrica*, July 1961, *29*, 315–35.

D. H. Pyle, "Observed Price Expectations and Interest Rates," *Rev. Econ. Statist.*, Aug. 1972, *54*, 275–80.

V. Tanzi, "Inflation, Indexation and Interest Income Taxation," *Banca Naz. Lavoro Quart. Rev.*, Mar. 1976, *29*, 54–76.

S. J. Turnovsky, "Some Empirical Evidence on the Formation of Price Expectations," *J. Amer. Statist. Assn.*, Dec. 1970, *65*, 1441–54.

Internal Revenue Service, *Statistics of Income-Annual, Individual Income Tax Returns*, Washington various issues.

Board of Governors of the Federal Reserve System, *Banking and Monetary Statistics, 1941–70*, Washington 1976, 693–707.

______, *Annual Statistical Digest, 1971–75*, Washington 1976, 121–24.

U.S. Council of Economic Advisors, *Economic Report of the President*, Washington 1977.

P17821

Oligopoly and Competition in Large Markets

By MASAHIRO OKUNO, ANDREW POSTLEWAITE, AND JOHN ROBERTS*

In this paper we study questions of oligopoly and competition in a general equilibrium framework. In particular, we consider the Nash equilibria of a model of noncooperative exchange in the context of a measure space of economic agents which incorporates both atoms, representing large traders or organized syndicates of traders, and a nonatomic continuum of infinitesimal individual traders.

Benyamin Shitovitz (1973, 1974) introduced this type of measure-theoretic model to study situations in which some but not all agents may have market power. Traditional general equilibrium treatments of such situations (see, for example, Kenneth Arrow and Frank Hahn, ch. 6) have been deficient in that they have simply assumed a priori that certain agents behave as price takers while others act noncompetitively, with no formal explanation being given as to why a particular agent should behave one way or the other. Shitovitz's approach represents an important contribution in pointing to an explicit formulation leading to such differences in behavior. In this paper we seek to explore the use of this type of model in studying issues of oligopoly in a general equilibrium framework. A specific focus of our work is in illuminating how either perfectly or imperfectly competitive behavior may emerge endogenously in this model, depending on the characteristics of the agent and his place in the economy.

Shitovitz's analysis concentrated on the core of the economy, that is, the set of allocations which no group of agents can improve upon by using only its own resources to achieve a distribution of commodities which each of its members prefers to the allocation in question. This solution concept has, of course, been widely applied in economics, and the equivalence between the core and competitive equilibria in the absence of large traders or syndicates is well known. (See Werner Hildenbrand for a presentation of these results.) Most of the succeeding work with Shitovitz's mixed measure-theoretic model has also been concerned with the core. (See, for example, Jean Jaskold-Gabszewicz and Jacques Drèze, Jaskold-Gabszewicz, Robert Aumann, Andrew Postlewaite and Robert Rosenthal, and Drèze, Gabszewicz, and Postlewaite.)

While Shitovitz's model would seem especially appropriate for studying oligopoly, he concentrated not so much on market power phenomena per se as on the possibility that all the core outcomes would still be competitive allocations despite the presence of atoms. Some of the results he obtained in studying this issue appear so counterintuitive as to seem to call into question the use of this model with atoms and a nonatomic ocean in studying oligopoly. For example, his Theorem B (1973) indicates that if there are two large traders or syndicates with the same endowment densities and preferences over consumption bundle densities, then the core and competitive allocations coincide, no matter what the relative sizes of the two traders. Thus, the presence of an arbitrarily small (but not infinitesimal) rival can be viewed as completely cancelling out the market power of a large trader who might represent all but a tiny fraction of the market's potential supply of a commodity.

Of course, if one finds such a result unsatisfying, one need not question the model of atoms and a continuum. Rather, one might object to the use of the core as the solution concept. Our work here indicates

*Okuno: University of Illinois and Yokohama National University; Postlewaite: University of Illinois; Roberts: Northwestern University. Part of this work was done while Postlewaite was a visiting faculty member at the University of California-San Diego. We wish to thank George Borts and an anonymous referee for comments. Roberts' research was supported by the National Science Foundation under grant SOC 76-20953, while Postlewaite's work was partially supported under grant SOC 77-27403.

that this is probably the right response. In place of the core, we take as our solution concept the Nash equilibrium of a noncooperative model of exchange. In this framework, we are able to obtain general equilibrium results concerning the qualitative properties of oligopoly which are both intuitively appealing and in accord with the insights offered by partial equilibrium analyses.

The model of exchange we use is derived from one introduced by Martin Shubik in which a good is exchanged for money. Each seller offers a quantity of the good for sale and receives a proportion of the total amount of money bid for the good by buyers equal to the proportion of the total amount of the good that he supplied. Similarly, a buyer bids an amount of money for the good and receives the fraction of the total supply that his bid is of the total amount bid. This model has been extended by Lloyd Shapley and Shubik to encompass many commodities, with one of the goods being used as commodity money.

With a finite number of agents, the Nash equilibria of the game defined by this rule of exchange are not typically Pareto optimal. One source of this inefficiency corresponds to imperfectly competitive behavior and is a central concern in this paper. However, the fixed initial supply of the commodity money can also lead to an inefficiency which may prevail even in large economies (see Jerry Jaynes, Okuno, and David Schmeidler for an analysis of this question). However, a variant of the model introduced by Elisha Pazner and Schmeidler eliminates this source of inefficiency by replacing the commodity money by units of account, and it is with this model that we will be concerned here.

Within this model, Postlewaite and Schmeidler have shown as the number of traders becomes large and each individual becomes small relative to the market, the inefficiency arising from imperfectly competitive behavior disappears asymptotically. Also Pradeep Dubey and Shapley have shown that the Nash and competitive equilibria coincide if there are an infinite number of traders, each of whom is negligible relative to the market. These results may be viewed as general equilibrium, pure exchange versions of the classical results on the competitiveness of the Cournot equilibrium in large economies. Our analysis focuses, on the other hand, not on the competitiveness of the outcomes of imperfectly competitive behavior, but rather on the nature of oligopoly where a small number of large traders face one another and a large mass of unorganized agents.

In the following section we specify the model we will be considering. We next present a simple example which illustrates the model. This example satisfies the assumptions of Shitovitz's Theorem B, but the Nash allocations (as opposed to the core) are not Pareto optimal, let alone competitive. In the succeeding section, we establish this property in general. We further show that the larger of two syndicates, each made up of the same type of traders, can be expected to restrict its sales and purchases more than does the smaller group. Finally we address two aspects of the incentives to form cartels. First, we show that any individual agent can gain by breaking from a cartel once it is formed. Second, we examine whether it is always advantageous for a group to collude or whether, instead, disadvantageous syndicates can exist in the sense that all members of the syndicate would gain from its breaking up.

Our aim in this paper is not to achieve the greatest possible generality. Rather, we seek to show how the framework introduced by Shitovitz, when combined with an appropriate model of noncooperative behavior, can lead to useful insights concerning oligopoly. For this reason, we will work with the simplest possible model, even though some of our arguments are applicable to more general situations.

I. The Model

We consider an exchange economy with two commodities in which the set of traders is indexed by the unit interval $T=[0,1]$ endowed with Lebesgue measure μ. Intuitively, we interpret $\mu(S)$ to be the proportion of the total set of traders belonging to the set S.

(Without further mention, we take all subsets of T and functions on T to be measurable with respect to the relevent measure. If atoms are present, this measure is not the Lebesgue measure itself, but is derived from it in a standard way.)

Each agent $t \in T$ is endowed with a strictly positive amount $w_i(t)$ of each good, $i=1,2$, and has preferences represented by a smooth, monotonic, strictly quasi-concave utility function $U^t(.,.)$.

To simplify further, we assume that there are but two types of agents, who are indexed respectively by the disjoint sets T^0 and T^1, $T^0 \cup T^1 = T$. The members of T^0 will be treated as unorganized traders throughout the discussion. Since our main interest in this paper is the effect of oligopolies or syndicates in markets, we will consider various different organizations of the traders in T^1. At the one extreme, T^1 may also consist of unorganized traders who act individually. Alternatively, T^1 might consist of a collection $(T^{11},\ldots,T^{1n})$ of syndicates, which are noninsignificant groups of traders who coordinate their actions within respective groups. This would represent a situation of oligopoly. As a third possibility there might be an unorganized portion of T^1 as well as one or more syndicates. The key point is that the members of a syndicate act in concert in deciding on the common trades they will all make, while those in the unorganized portion of the market act individually.[1]

Trade takes place through a market in which the two goods are exchanged against one another. Each trader sends some amount of one or the other commodity to market. These supplies are aggregated, and a price ratio is obtained as the ratio of these aggregate supplies. Each trader sending good 1 to the market then receives the amount of good 2 his supply will purchase at the effective price, and similarly for suppliers of good 2. For an unorganized trader, his supply decisions have a completely negligible effect on the market aggregates and thus on the price ratio. Consequently, he faces a budget constraint that coincides with that of the usual Walrasian price taker. However, organized groups of traders acting in concert exert a nonnegligible impact on prices: in general, an increase in the supply of a good by a syndicate will lower the relative price of the good. Therefore, the budget set facing an organized trader is characterized by a strictly concave curve rather than by a straight budget line. Agents recognize this phenomenon. Consequently, when organized as a syndicate they adopt behavior which is not perfectly competitive although, when unorganized, the same agents with the same characteristics and motivation would be price takers.

More specifically, let $s: T \to R^2_+$ be an integrable function associating with each trader a strategy, that is, $s_i(t)$ is the amount of the ith good that trader t sends to the market. We will call such a function a supply profile and denote it $(s_1(\cdot), s_2(\cdot))$. A syndicate $S \subset T^1$ is a subset of T^1 of positive measure whose members coordinate their supply decisions. By assumption, if t and t' both belong to some syndicate S, then $s_i(t) = s_i(t') = s_i$, $i=1,2$.

If an unorganized trader t supplies (σ_1, σ_2) while the supply profile of the economy is $(s_1(\cdot), s_2(\cdot))$, this unorganized trader will receive in return a density of

$$(1) \quad b_1^t(\sigma, s(\cdot)) \equiv \sigma_2 \int_T s_1(\tau)\, d\mu \Big/ \int_T s_2(\tau)\, d\mu$$

of the first good and

$$(2) \quad b_2^t(\sigma, s(\cdot)) \equiv \sigma_1 \int_T s_2(\tau)\, d\mu \Big/ \int_T s_1(\tau)\, d\mu$$

of the second good. A trader t who belongs to a syndicate S of measure (size) μ^S, each member of which is supplying (σ_1, σ_2), receives

$$(3) \quad b_1^t(\sigma, s(\cdot)) \equiv \sigma_2 \left[\int_{T \sim S} s_1(\tau)\, d\mu + \sigma_1 \mu^S \right] \Bigg/ \left[\int_{T \sim S} s_2(\tau)\, d\mu + \sigma_2 \mu^S \right]$$

[1] No nontrivial equilibrium exists in the model we are considering if all the traders in T^1 are organized into a single syndicate. Thus, we are unable to model pure monopoly without a competitive fringe in a useful way in this setup.

and

$$(4) \quad b_1^t(\sigma,s(\cdot)) \equiv \sigma_1\left[\int_{T\sim S} s_2(\tau)\,d\mu + \sigma_2\,\mu^S\right] \Big/ \left[\int_{T\sim S} s_1(\tau)\,d\mu + \sigma_1\,\mu^S\right].$$

We will assume that an agent never supplies both goods to the market, i.e., that $s_1(t)\cdot s_2(t)=0$ for all t. Thus, either b_1^t or b_2^t will be zero for each t at any profile.

A Nash equilibrium supply profile, given the organization of T^1, is a profile $(s_1(\cdot)), s_2(\cdot))$ such that each unorganized trader t is maximizing his utility with respect to his choice $(s_1(t),s_2(t))$ and each syndicate S^j is choosing a common trade $(s_1(t),s_2(t))=(s_1^j,s_2^j)$ for each member of the syndicate such that the common utility function of all members of S^j is maximized by this choice. More formally, let the sets Σ^t be defined by $\Sigma^t=[s\in R_+^2 | s\leqq w(t)]$. Define the pay-off function h^t for any trader, given his choice $\sigma\in\Sigma^t$ and given s, by

$$(5) \quad h^t(\sigma,s)=U^t(w(t)+b^t(\sigma,s(\cdot))-s(t))$$

Then, given the organization of the economy into syndicates and unorganized traders, a Nash equilibrium is a supply profile $(s_1(\cdot),s_2(\cdot))$ such that for each unorganized t,

$$(6) \qquad h^t(s(t),s)\geqslant h^t(\sigma,s)$$

for all $\sigma\in\Sigma^t$ and such that for each syndicate $S\subset T^1$ there does not exist σ' such that $\sigma'\in\Sigma^t$, $t\in S$, and

$$(7) \qquad h^t(\sigma',s)>h^t(s(t),s)$$

for the traders in S. (Recall Σ^t and h^t are constant on $S\subset T^1$.)

It should be noted that if $s_i(t)\equiv 0$ for all t, then this game is not well defined. By introducing an arbitrary market allocation scheme to deal with such situations, one can avoid this problem. However, this procedure usually results in the creation of trivial Nash equilibria in which no trade occurs ($s_1(t)=s_2(t)=0$ for all t). Note that such an equilibrium would also result if the initial endowment were Pareto optimal, so that no mutually advantageous trade is possible. Throughout, we will ignore such no-trade equilibria.

II. An Example

At this point, an example may be useful. Suppose $\mu(T^0)=1/2$ and that T^1 is organized into two syndicates, S^1 and S^2, where $\mu(S^1)=\mu(S^2)=1/4$. The agents in T^1 have straight indifference curves with slope $dx_2/dx_1=-2$, while those in T^0 have indifference curves whose slope is $-3/2$ for $x_2>3x_1$ and 0 for $x_2<3x_1$. The endowments densities are (2,2) for all agents.

The unique competitive equilibrium in this economy gives consumption densities of (2/3,4) to each agent in T^0 and of (10/3,0) to each in T^1. However, this allocation does not correspond to a Nash equilibrium, while the allocation giving (1,3) to the agents in T^0 and (3,1) to all in T^1 is the unique Nash equilibrium allocation. To verify that the competitive allocation is not Nash, note that a simple calculation shows that the budget set facing a member of S^1 near the competitive outcome is given by

$$(x_1,x_2)=\left(2+\frac{8\sigma_2}{3(2+\sigma_2)},2-\sigma_2\right)$$

where σ_2 is the per capita amount of good 2 supplied by S^1 and S^1 supplies none of the first good. The slope of this curve is $dx_2/dx_1=-9(2+\sigma_2)^2/48$, which at the competitive supply of $\sigma_2=2$ is -3. But the slope of the indifference curve is -2, so this point involves too great a supply to be maximizing for the members of S^1. Thus, it cannot be a Nash equilibrium. To verify that the claimed allocation is a Nash equilibrium, consider the options open to the first syndicate. If it considers supplying an amount σ_2 in per capita terms of good 2, while S^2 continues to supply 1 unit per capita of good 2 and T^0 supplies 1 unit per capita of the first good, then the members of S^1 will

receive in return

$$\frac{\sigma_2(1/2)1}{(1/4)1+(1/4)\sigma_2}=\frac{2\sigma_2}{1+\sigma_2} \tag{8}$$

of good 1 per capita. If on the other hand S^1 considers supplying σ_1 per capita of good 1, each member will receive in return

$$\frac{\sigma_1(1/4)1}{(1/2)1+1/4\sigma_1}=\frac{\sigma_1}{2+\sigma_1} \tag{9}$$

per capita of good 2. Given the fixed supplies from T^0 and S^2, these exchange possibilities give rise to consumption densities of

(10)
$$(x_1,x_2)=\left(2+\frac{2\sigma_2}{1+\sigma_2}-\sigma_1, 2+\frac{\sigma_1}{2+\sigma_1}-\sigma_2\right)$$

where either σ_1 or σ_2 (or both) is zero. The corresponding attainable set is diagramed in Figure 1. It is readily verified that the intercepts are as indicated and that the frontier is given by the differentiable, strictly concave function $x_2=(3x_1-10)/(x_1-4)$. Thus, the marginal cost of good 1 in terms of good 2 is $-2/(x_1-4)^2$. At the specified point, (3, 1), this is equal to -2, the *MRS* of any agent in T^1. Thus, as claimed, the consumption density (3, 1) is optimal for S^1. Of course, the same is true for S^2.

Now consider any trader $t\in T^0$. Such a trader faces a Walrasian budget constraint passing through his initial endowment and having slope of $dx_2/dx_1=-p=-\int s_2/\int s_1$. Given the supply profile considered here, the price p is $(1/2)/(1/2)=1$. With the specified preferences for the agents in T^0, the optimal consumption is, as claimed, (1, 3). Thus, the specified allocation is a Nash equilibrium. Its uniqueness can be checked in a straightforward manner.

This Nash equilibrium is not competitive: indeed, it is not even Pareto optimal, since it involves too low a level of trade. In contrast, if the core were our solution concept, then by Shitovitz's Theorem B, any equilibrium allocations would be competitive. Thus, the noncooperative model leads to sharply differing results from the cooperative model of exchange. Moreover, the nature of the equilibrium outcomes is consistent with the intuition suggested by partial equilibrium analysis. Those agents with market power take into account the effect of an increase in supply on their part on the terms of trade. Thus their maximizing trades are not the Walrasian trades they would make if they had no impact, but instead are smaller.

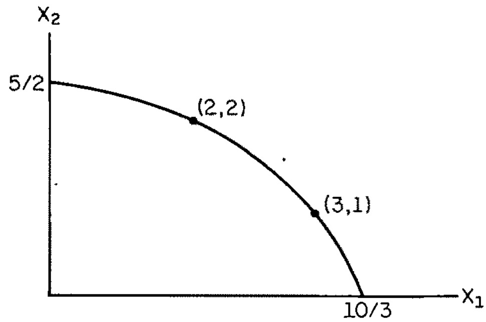


FIGURE 1

III. Oligopolistic Equilibria and the Incentives to Organize

It is clear that if some agents are organized the equilibria of this model cannot in general be Pareto optimal. Maximizing behavior requires that an agent's marginal rate of substitution equal the marginal rate at which he can trade off one good for the other. For unorganized agents, this marginal price equals the average price, while the two differ for organized traders. Thus, the marginal rates of substitution will differ between agents at equilibrium and, for interior solutions, this implies inefficiency. In this section we obtain further qualitative properties of the Nash equilibria.

At a competitive equilibrium, all agents with given tastes and endowments receive the same consumption (or, if preferences are not strictly convex, at least bundles that are indifferent). This property does not hold at the Nash equilibrium, as we now show.

PROPOSITION: *Let S^1 and S^2 be two syndicates in T^1 of positive but differing sizes and let $s(\cdot)$ be a nontrivial Nash equilibrium supply profile resulting in a strictly positive*

per capita consumption vector for the members of each of these syndicates. Then $x^1 \neq x^2$, where x^i is the equilibrium consumption vector for $t \in S^i$.

PROOF:

Let s be a Nash equilibrium and, to obtain a contradiction, suppose that $x^1 = x^2$. Since $x^1 = x^2$, we must have $s^1 = s^2$, where $s^i = (s_1^i, s_2^i)$ is the per capita supply of syndicate i. We will further assume without loss of generality that $s_1^1 = s_1^2 = 0$, that is, the organized traders are suppliers of the second good and purchasers of the first.

Now, if the syndicate S^2 supplies σ_2^2 per capita to the market while the other agents supply the amounts from the equilibrium profile, then it would obtain the amount

$$(11) \qquad b_1^2 = \frac{\sigma_2^2 \int s_1(t)}{\mu^2 \sigma_2^2 + \mu^1 s_2^1}$$

of good 1 per capita in return. Therefore, the marginal price ratio between good 1 and good 2 for S^2 is

$$(12) \qquad \frac{-db_2^2}{db_1^2} = \left[\frac{\mu^2 \sigma_2^2 + \mu^1 s_2^1}{\int_{\sim(S^1 \cup S^2)} s_1(t)} \right] \cdot \left(\frac{\mu^2 \sigma_2^2 + \mu^1 s_2^1}{\mu^1 s_2^1} \right)$$

At the Nash equilibrium strategy, $\sigma_2^2 = s_2^2$, so

$$(13) \qquad \frac{-db_2^2}{db_1^2} = p \left(\frac{\mu^2 s_2^2 + \mu^1 s_2^1}{\mu^1 s_2^1} \right)$$

where $$p = \frac{\mu^2 s_2^2 + \mu^1 s_2^1}{\int_{\sim(S^1 \cup S^2)} s_1(t)}$$

is the average price of good 1 in terms of good 2 at equilibrium.

Since the Nash equilibrium results in maximization of per capita utility, we have (assuming $x^2 > 0$) that the marginal rate of substitution for members of S^2 must equal the marginal price. Thus, $MRS^2 = p(\mu^2 s_2^2 + \mu^1 s_2^1)/\mu^1 s_2^1$. By the same reasoning, $MRS^1 = p(\mu^1 s_2^1 + \mu^2 s_2^2)/\mu^2 s_2^2$.

By assumption, the per capita consumption allocations are the same to the syndicates, and since the utility functions are the same, we must have $MRS^1 = MRS^2$ at equilibrium. Thus, $\mu^1 s_2^1 = \mu^2 s_2^2$. But this cannot happen, given $s^1 = s^2$ and $\mu^1 \neq \mu^2$. Thus, assuming an interior solution for each syndicate, the agents in the two unequal-sized syndicates must be treated differently. It is perhaps worth noting that this proposition is valid in much more general situations than those specified here.

It seems intuitively appealing that the larger of two syndicates should restrict its per capita supply more than the smaller one, since its supply has a larger marginal impact on price. In the case that the good being sold by the members of T^1 shows nonnegative income effects under the T^1 preferences, this is in fact the case.[2] At a Nash equilibrium, the average price ratio is the same for all agents, so both syndicates' trades lie along the line with this slope through the common endowment. Given nonnegative income effects, the MRS is strictly decreasing along this budget line as x_1 increases. Thus, MRS^i/s_2^i is also decreasing if we assume that the agents in T^1 are selling good 2. But in this circumstance, $\mu^1/\mu^2 = (MRS^1/s_2^1)/(MRS^2/s_2^2)$ at equilibrium, as is seen from the argument above. Thus, if $\mu^1 > \mu^2$, we have $s_2^1 < s_2^2$ at equilibrium. Similarly, if the syndicates are selling good 1, then $s_1^1 < s_1^2$. Thus, the larger organized group will make a smaller trade per capita. This further implies that the larger group will achieve a lower per capita equilibrium level of utility.

[2]It is simple to show that at equilibrium no two members of T^1 can be on opposite sides of the market. For if t and t' are in T^1 and t is a net seller of good 1, say, while t' is a net seller of good 2, then at equilibrium the MRS for t is strictly less than that for t', while t consumes less of good 1 than does t'. This means that, under the common preferences, both goods show negative income effects, which is impossible under our assumptions.

There is an interesting sidelight to the above calculations. From (13) we see that $db_1^2/d\sigma_2^1$ (which can be thought of as marginal revenue for S^2) is equal to

$$(14)\quad \frac{(1/p)\mu^1 s_2^1}{\mu^1 s_2^1 + \mu^2 s_2^2} = \frac{1}{p}\left[1 - \frac{\mu^2 s_2^2}{\mu^1 s_2^1 + \mu^2 s_2^2}\right]$$

$$= \frac{1}{p}\left[1 - \text{the market share of } S^2\right]$$

But $1/p$ is just the price of good 2 in terms of good 1. Thus we see that marginal revenue of a syndicate can be written in terms of the price and its "degree of market control." The expression is reminiscent of the expession derived by partial equilibrium techniques for marginal revenue for a firm, $MR = p(1 - 1/\eta)$, where η is the absolute value of the elasticity of demand. The term $1/\eta$ is, of course, equal to Lerner's index of monopoly power when profit is maximized.

The fact that an agent's equilibrium consumption depends on whether or not he belongs to an organized group and, if he does, on the size of the group to which he belongs raises the issue of stability of syndicates. We consider two aspects of this question. First, we look at the incentives facing a trader to join or break from an existing syndicate individually, and second we consider whether a group of agents will collectively gain forming a syndicate and withholding supply. As might be expected, there is reason for an individual to act as a free rider and break from a syndicate as long as he expects the cartel to stay otherwise intact. However, we also show that the formation of a syndicate may prove disadvantageous to all of its members in that they all receive a less preferred consumption bundle when organized than they would if they *all* acted competitively. We then present a sufficient condition for this apparent anomaly not to arise.

A syndicate typically may be expected to withhold supply, trading less than it would under competitive behavior at the prevailing (average) exchange ratio, since it faces a marginal price for the good it is purchasing that is higher than the average price at

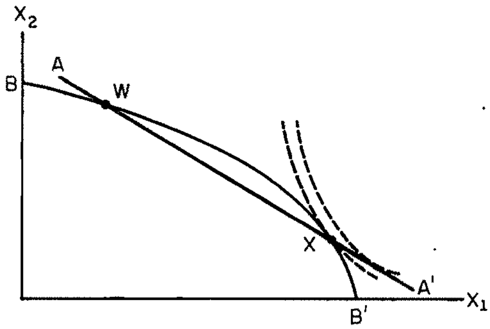


FIGURE 2

equilibrium. This is illustrated in Figure 2, where w is the endowment and x is the consumption chosen by the syndicate in per capita terms.

Now, an individually infinitesimal, unorganized trader outside the syndicate faces the achievable set given by the average price line AA', as opposed to that faced by the syndicate, which is given by BB'. The former set clearly contains points preferred to x. Moreover, if the syndicate will continue in existence, the defection of a single agent will not alter the average price ratio, so these preferred points would be attainable to him. Thus, any member of any syndicate has an incentive individually to leave the syndicate if he expects it to remain in existence despite his defection.

This result may, however, overstate the extent of cartel instability in finite economics, since in actuality no trader is truly negligible. Thus, he cannot automatically assume that the syndicate will not dissolve, and, as well, his adopting competitive behavior will affect the price ratio. Both of these effects tend to lessen the expected gains to departing from the syndicate, and to impart some greater stability to syndicates. For a similar analysis in a somewhat different framework, see Postlewaite and John Roberts.

However, even if syndicates are unstable in this sense, one might expect that there would still be an incentive to organize syndicates. The instability noted above can be explained in terms of the individual agent acting as a free rider, gaining the benefits of the syndicate's restriction of supply without

contributing to the burden of maintaining this restriction. However, given a group of agents with identical tastes and endowments, one might expect that they would collectively gain from collusion, even if, once the group is formed, any single individual might gain from departing from it.

In fact, the situation is much more complicated than might seem. We illustrate this by considering two economies, each of which may be organized in differing ways.

The first economy is that given in the earlier example, in which all agents have an endowment (2,2), half the agents (those in T^0) have preferences with $MRS=3/2$ for $x_2>3x_1$ and 0 for $3x_1>x_2$, and the other half (T^1) have straight line indifferences curves with slope -2. As noted earlier, the competitive equilibrium in this economy, which is also the Nash equilibrium when all agents are unorganized, yields consumption vectors of (2/3,4) for traders in T^0 and (10/3,0) for those in T^1. Also as noted earlier, if all the traders in T^1 organize into two equal-sized syndicates, the per capita outcome is (1,3) for the agents in T^0 and (3,1) for each member of each syndicate. In this case, the T^1 agents gain from being organized, since they prefer (3,1) to (10/3,0).

The second example is identical to the first, except that for agents in T^0 the MRS is 6/5 for $x_2>3x_1$, rather than 3/2. In this economy, the competitive outcome gives (1/3,4) to the agents in T^0 and (11/3,0) to those in T^1, while the outcome when the T^1 agents are organized in two equal-sized syndicates is as in the first example, (3,1) for $t\in T^1$. In this case, the organized agents are all worse off as a result of being organized, since they prefer (11/3,0) to (3,1).

Another possible organization of these economies shows further complexities. Suppose that in each economy half the T^1 agents are organized as a syndicate while the other half and the agents in T^0 are unorganized. In the first economy this results in a Nash equilibrium yielding (10/9,10/3) to the agents in T^0, (10/3,0) to the unorganized agents in T^1 and (22/9, 4/3) to the organized agents. In the second economy these vectors are approximately (.62,3.67), (3.67,0), and (3.11,.67) respectively. Using the utility function $u^1(x_1,x_2)=2x_1+x_2$ and taking the T^1 agents to consist of two equal-sized groups, either of which can be organized or unorganized, we can represent the outcomes to the T^1 agents in tabular form as shown in Table 1. Thus, in the first economy, if the agents represented by the rows are organized while those represented by the columns are unorganized, the outcome gives utility 6.22 to agents in the first group and 6.67 for those in the second.

TABLE 1—PER CAPITA UTILITIES FOR MEMBERS OF T^1

Economy 1

Group 2 \ Group 1	Organized	Unorganized
Organized	7 / 7	6.67 / 6.22
Unorganized	6.22 / 6.67	6.67 / 6.67

Economy 2

Group 2 \ Group 1	Organized	Unorganized
Organized	7 / 7	7.33 / 6.89
Unorganized	6.89 / 7.33	7.33 / 7.33

In the first economy, the best outcome from the point of view of a T^1 agent is when both groups are organized. If he and his group are unorganized, it is a matter of indifference to him whether the other group organizes or not. The worst outcome is for him to be in an organized group while the other group's members are acting competitively. Thus, starting from a competitive organization, one would not expect either group to form unilaterally. If one starts from a position of full cartelization, this too is stable. And one might expect an

organized syndicate facing a competitive fringe would dissolve if it were able to predict the outcome that would thereby result.

In the second economy the ranking of the outcomes differs. Here the T^1 agents are best off if they are completely unorganized. Any agent in one group is, however, indifferent as to whether the other group organizes or not, given that his group is unorganized. Either of these situations is preferred by him to the situation in which both syndicates are organized, while the worst outcome again results from being a member of a organized group while the other group behaves competitively. In this economy, then, the cartelized outcome is unstable, while the competitive outcome is stable. Any movement towards organization hurts those attempting it.

This matter of the formation of a syndicate hurting the agents involved in the syndicate is reminiscient of the phenomenon of disadvantageous syndicates noted in connection with the core by Aumann and investigated by several other authors since. Generally the formation of a syndicate will enlarge the set of core allocations since the coalitions involving only some of the members of the syndicate are no longer available for blocking. One then says that a syndicate is disadvantageous if the new allocations entering the core are worse for the members of the syndicate than the allocations originally there.

A major focus of research on the issue of disadvantageous syndicates has been to explain their origin and existence in economic terms. In the context of our noncooperative model, there seems to be an obvious explanation. In particular, an agent (or syndicate) in deciding on his best action takes the bids forthcoming from the rest of the economy as fixed, that is, he assumes demand for the good he supplies is of unitary elasticity. This is, in general, clearly not a correct expectation and a syndicate's impact is such that this misperception can become important.

Suppose that, in fact, the offer curve from the unorganized portion of the economy actually did have the assumed shape, at least in a region of the Nash equilibrium.

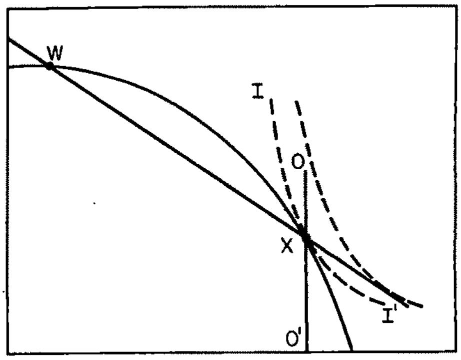


FIGURE 3

This is illustrated in Figure 3, which may be interpreted in terms of one side of the market consisting either of a single syndicate and an unorganized group of equal size or of two equal-size syndicates. Here W is the endowment point, X is the Nash equilibrium and the line connecting them gives the average price ratio. The curve II' is the indifference curve of a member of the syndicate through the equilibrium, which necessarily is steeper than the price line since the marginal price exceeds the average for the syndicate. The vertical line $00'$ is the offer curve, in per capita terms, of the unorganized part of the economy.

The intersection of the competitive offer curve of the syndicate with $00'$ would give a competitive equilibrium. Since this offer curve must meet II' below and to the right of X, must pass through W, and is continuous, it is clear that it must intersect $00'$ below II'. Thus, there will be a competitive allocation which is worse for the syndicate than the Nash allocation.

In fact, as long as the T^0 offer curve passes through X and never doubles back so as to be above II' again below and to the right of X, there will be such a competitive allocation. In this regard, it is useful to compare the two examples given earlier: this condition is met in the first example when both groups are organized, and the organized outcome is better than the competitive one; it is not met in the second, and the syndicates are disadvantageous. Of course, this condition on offer curves near

the equilibrium is unsatisfactory, since it involves the equilibrium rather than being stated directly on the primitive data of tastes and endowments. We leave it as an interesting but open question as to whether conditions on preferences and endowments can be devised to insure that organizing is beneficial to the traders involved.

IV. Conclusions

We have examined a very simple general equilibrium model of exchange in which some agents, by acting collectively, have monopoly power, while others, acting individually, are competitive price takers. The results and examples we present indicate to us that this model is a useful one for examination of oligopoly. In particular, the results on the qualititative nature of the equilibria—that they are nonoptimal and that, in the normal case, large groups restrict supply more than small ones—would seem to be useful results. The issue of cartel stability which we have discussed is also significant and worthy of further work. In particular, our results indicate that cartelization need not always be profitable for a given group of agents, since we have shown that every agent in a syndicate may be better off if the syndicate dissolves and its members act competitively. Further, this phenomenon can occur whether or not the other agents on the same side of the market as the syndicate are colluding among themselves. This raises the difficult but fascinating issue of building a model in which one can explain endogenously which cartels will emerge.[3]

[3]For an interesting contribution in this direction using the von Neumann-Morgenstern solution, see Sergiu Hart.

REFERENCES

Kenneth J. Arrow and Frank H. Hahn, *General Competitive Analysis*, San Francisco 1971.

R. H. Aumann, "Disadvantageous Monopolies," *J. Econ. Theory*, Feb. 1973, *6*, 1–11.

J. Drèze, J. Jaskold-Gabszewicz, and A. Postlewaite, "Disadvantageous Monopolies and Disadvantageous Endowments," *J. Econ. Theory*, Oct. 1977, *16*, 116–21.

P. Dubey and L. Shapely, "Noncooperative Exchange with a Continuum of Traders," Cowles Foundation disc. paper no. 447, Yale Univ. 1977.

S. Hart, "Formation of Cartels in Large Markets," *J. Econ. Theory*, Apr. 1974, *7*, 453–466.

Werner Hildenbrand, *Core and Equilibria of a Large Economy*, Princeton, 1974.

J. Jaskold-Gabszewicz, "Asymmetric Duopoly and the Core," *J. Econ. Theory*, Feb. 1977, *14*, 172–179.

______ **and J. H. Drèze**, "Syndicates of Traders in an Exchange Economy," in H. W. Kuhn and G. P. Szego, eds., *Differential Games and Related Topics*, Amsterdam 1971, 399–414.

J. Jaynes, M. Okuno, and D. Schmeidler, "Efficiency in an Atomless Economy with Fiat Money," *Int. Econ. Rev.*, Feb. 1978, *19*, 149–156.

E. Pazner and D. Schmeidler, "Non-Walrasian Equilibria and Arrow-Debreu Economies," mimeo., Univ. Illinois.

A. Postlewaite and J. Roberts, "A Note on the Stability of Large Cartels," *Econometrica*, Nov. 1977, *45*, 1877–78.

______ **and R. Rosenthal**, "Disadvantageous Syndicates," *J. Econ. Theory*, Nov. 1974, *9*, 324–326.

______ **and D. Schmeidler**, "Approximate Efficiency of Non-Walrasian Nash Equilibria," *Econometrica*, Jan. 1978, *46*, 127–36.

L. Shapley and M. Shubik, "Models of Noncooperative Exchange," unpublished paper.

B. Shitovitz, "Oligopoly in Markets with a Continuum of Traders," *Econometrica*, May 1973, *41*, 467–502.

______, "On Some Problems Arising in Markets with Some Large Traders and a Continuum of Small Traders," *J. Econ. Theory*, Aug. 1974, *8*, 458–70.

M. Shubik, "Commodity Money, Oligopoly, Credit and Bankruptcy in a General Equilibrium Model," *Western Econ. J.*, Mar. 1972, *10*, 24–38.

Price Inflation, Portfolio Choice, and Nominal Interest Rates

By BENJAMIN M. FRIEDMAN*

The belief that expectations of future price inflation tend to result in higher nominal interest rates—the proposition which Irving Fisher, in a perhaps more optimistic era, labeled "appreciation and interest"[1]—is now commonplace among both economists and financial market participants. The rise of nominal interest rates in the United States to record high levels in 1974 and again in 1979, just when the *U.S.* economy was undergoing its first experiences of peaçetime double-digit inflation, demonstrates this relationship at even the most unsophisticated eyeball level. Many economists have also undertaken statistical investigations of this relationship, using a variety of devices to obviate the analytical difficulties due to the unobservability of the relevant expectations. With its immediate implications for real yields, the Fisher relationship is central to the classic questions confronting monetary theory and policy.

Many important aspects of the expected price inflation/nominal interest rate relationship remain undetermined, however. Even at the most fundamental level, little is known about the exact nature of the process by which economic behavior causes nominal interest rates to respond to price expectations in the specified way. Fisher himself was curiously sketchy in his "interpretation" of the relationship,[2] and most subsequent writers on the subject have followed his lead in seeking more to document and quantify the relationship than to investigate in any precise way the underlying economic behavior which causes it.[3]

The object of this paper is to explore in some detail one of the possible sources of the Fisher relationship—in particular, the portfolio behavior of lenders. The starting point for this analysis is the simple truism that any factor hypothesized to affect interest rates must affect some lender's demand

*Harvard University. I am grateful to Orlin Grabbe, David Jones and Vance Roley for research assistance, and to them as well as James Duesenberry, Stanley Fischer, Robert King, John Lintner, Thomas Mayer, Allan Meltzer, Franco Modigliani, John Rutledge, Jerome Stein, James Tobin, and an anonymous referee for helpful discussions and comments on a previous draft of this paper. I also gratefully acknowledge the research support of the National Science Foundation (grant APR77-14160) and the National Bureau of Economic Research.

[1]See especially Fisher (1896; 1907, ch. 5; 1930, ch. 19).

[2]See Fisher (1930, pp. 438–42). Somewhat astonishingly to the modern reader, Fisher's suggested interpretation followed Knut Wicksell in noting that higher prices usually meant a greater nominal volume of trade, which in turn increased the demand for money, and hence increased nominal interest rates for given bank reserves. What is surprising about this interpretation is that, as rendered by Fisher, it has nothing whatever to do with expectations. In addition, the association of higher prices with a greater nominal transactions volume is necessarily valid only if the source of the inflation is a demand shock to the economy; under a supply shock prices rise but real transactions volume falls, so that the change in nominal transactions volume remains ambiguous in general. At least in his early work, Fisher also hypothesized asymmetrical behavior between lenders and borrowers (see 1896, pp. 75–78), and the useful discussion in John Rutledge (1977).

[3]Moreover, perhaps in part because of his imprecision about the underlying behavior, there exists today substantial misinformation about Fisher's views on the resulting relationship. For example, on the short-run invariance of the real rate with respect to inflation, "...in actual practice...the appreciation or depreciation of the monetary standard does produce a real effect on the rate of interest.... This effect, in times of great changes in the purchasing power of money, is by far the greatest of all effects on the rate of interest" (1930, p. 493). On constancy of the real rate for given inflation, "...there are...so many other causes affecting the rate of interest besides changes in the price level" (1930, p. 411). These views clearly contradict such interpretations as the "Fisherian" proposition tested by Eugene Fama (1975).

for loans, or some borrower's supply of loans, or both.[4] If expectations of price inflation raise nominal interest rates, they must do so by reducing lenders' willingness to lend and/or increasing borrowers' willingness to borrow at a given nominal yield. The responses to such expectations on the part of lenders and borrowers could logically involve not only their portfolio behavior (the composition of their assets held and liabilities issued), but also their saving and investment behavior (the amounts of their assets held and liabilities issued).[5]

This paper investigates the role of lenders' portfolio behavior in the relationship between price expectations and nominal interest rates, using behavioral equations directly explaining lenders' willingness to enter into long-term fixed-income loan contracts.[6] To anticipate, the conclusion of the paper's partial-equilibrium analysis is that lenders' portfolio behavior is an important component of the economic process generating the Fisher relationship. Results based on *U.S.* data indicate that, with all other aspects of economic behavior held unchanged, lenders' portfolio behavior would cause the equilibrium level of nominal bond yields to rise by 0.65 percent for each 1 percent of expected price inflation.

Section I develops equations representing lenders' behavior in the market that many previous researchers have also chosen to reflect most clearly the Fisher relationship —in particular, the market for long-term fixed-interest loans (bonds)—and Section II presents estimation results for these equations based on *U.S.* data for six major categories of lenders. Section III presents partial-equilibrium simulation results showing the implications of lenders' portfolio behavior, as represented by these estimated equations, for the relationship between expected price inflation and nominal yields. Section IV briefly summarizes the paper's conclusions.

I. A Model of Lenders' Demand for Loans

It is well known that, for risk-averse investors maximizing the expected utility of either terminal wealth or portfolio rate of return, the optimal portfolio allocation depends not only on the means but also on the higher moments of the distributions describing the expected holding-period rates of return on each individual asset available for portfolio investment. Utility functions exhibiting constant relative risk aversion, together with joint normal (or lognormal) asset return distributions, generate optimal single period portfolio allocations that are linear homogeneous in wealth and linear in the expected asset returns.[7] Further linearization, to separate the effects of the expected returns from the associated variance-covariance structure, yields the investor's desired portfolio allocation in the form

$$(1)\qquad \frac{A_{it}^*}{W_t} = \sum_k^N \beta_{ik} r_{kt}^e + \sum_k^N \gamma_{ik} v_{kt} + \sum_k^N \sum_{j\neq k}^N \delta_{ikj} c_{kjt} + \pi_i \qquad i=1,\ldots,N$$

[4]In well-developed financial markets, of course, the relevant group of "lenders" includes not only those who make primary loans directly to borrowers, but also those who may under certain circumstances be willing to acquire debt securities in a secondary market.

[5]In simple consumption-loan models there is typically no meaningful distinction between saving behavior and portfolio behavior in the conventional sense as meant here; see, for example, Paul Samuelson (1958). Even in models in which the two kinds of behavior have distinct meanings, they are not in general independent; see, for example, Fama (1970), Robert Merton, and Samuelson (1969). Hence the portfolio behavior analyzed in Section I explicitly relies on a single period horizon.

[6]See my 1978 paper for an analogous treatment of the role of borrowers' portfolio behavior.

[7]See my 1979 paper with V. Vance Roley for a review of the alternative sets of assumptions which permit the derivation of asset demand functions that satisfy the wealth homogeneity property and that are linear in expected asset yields; Frank De Leeuw and Milton Friedman also provided discussions of the rationale behind the homogeneity constraint. A particular advantage of the homogeneous form, in the context of this paper's concern with price inflation, is that rising dollar magnitudes per se do not affect portfolio allocation.

where

$A^*_{it}, i=1,\ldots,N=$ the investor's desired holding of the ith asset at time period t$(\Sigma_i A^*_{it} = W_t)$

$W_t=$ the investor's total portfolio size (wealth) at time period t

$r^e_{kt}, k=1,\ldots,N=$ the expected value of the holding-period yield on the kth asset at time period t

$v_{kt}, k=1,\ldots,N=$ the variance associated with r^e_{kt}

$c_{kjt}, k,j=1,\ldots,N=$ the covariance associated with r^e_{kt} and r^e_{jt}

and the β_{ik}, γ_{ik}, δ_{ikj}, and π_i are fixed coefficients which satisfy $\Sigma_i\beta_{ik}=\Sigma_i\gamma_{ik}=0$ for all k, $\Sigma_i\delta_{ikj}=0$ for all k and j, and $\Sigma_i\pi_i=1$. On the assumption of universal substitutability, the β_{ik} and γ_{ik} also satisfy $\beta_{ik}>0>\gamma_{ik}, i=k$, and $\beta_{ik}<0<\gamma_{ik}, i\neq k$. The expression for the investor's desired holding of loans (L^*_t), for example, is (without the covariances)

$$(2)\quad \frac{L^*_t}{W_t}=\beta_{LL}r^e_{Lt}+\sum_{k\neq L}^{N}\beta_{Lk}r^e_{kt}+\gamma_{LL}v_{Lt} + \sum_{k\neq L}^{N}\gamma_{Lk}v_{kt}+\pi_L$$

where the coefficients of the holding period yield expectations and variances satisfy $\beta_{LL}>0>\gamma_{LL}$ and $\beta_{Lk}<0<\gamma_{Lk}, k\neq L$.

What is the role of expected price inflation within a portfolio choice framework like (1) and (2)? If real assets constitute a plausible asset for portfolio investment, then the effect of price expectations here is straightforward. The increase in goods prices is then simply the holding-period return on investment in such real assets, and the associated expectation r^e_P is an argument of (2). The associated coefficient β_{LP} is negative so that, for all other things equal, greater expectations of price inflation reduce lenders' demand for loans. In practice, however, the available opportunities for such portfolio substitutions involving consumption or other real commodities are usually extremely limited; purchasing the Consumer Price Index (*CPI*) basket of goods is not a feasible portfolio alternative.

For expected price inflation to influence lenders' portfolio behavior under limited portfolio substitution possibilities, therefore, inflation expectations must differentially affect the expected holding-period returns (or variances or covariances) on those assets that do constitute plausible portfolio choices. For example, expected inflation may increase the expected return to equities relative to expected returns on money and loans (of all maturities) which have fixed nominal yields. Alternatively, investors may expect the future price inflation to bring higher future nominal interest rates—because of eventual effects on saving-investment behavior, or perhaps even because of a monetary policy response to inflation per se—thereby reducing the expected one-period return to holding a long-term loan relative to a short-term loan (or, equivalently, reducing the expected multiperiod return on a long-term loan relative to that on a series of short-term loans). Even for an investor who cannot practically invest in real assets, therefore, the demand-for-loans expression (2) may depend on expectations of price inflation via their influence on the expected asset returns r^e_k.[8] Moreover, if the investor associates uncertainty of asset returns with uncertainty of price inflation, then the variance (and relevant covariances) of price expectations will also be a determinant of the demand for loans.

[8]While such capital-loss-avoiding behavior indicates a positive connection between expected price inflation and long-term interest rates, this argument does not carry over to short-term interest rates. Hence it may be more accurate to refer to such behavior as producing a "term-structure effect" rather than a "Fisher effect," but it is difficult to support any such sharp distinction on the basis of Fisher's own work; see fn. 2.

The general portfolio allocation model (1), as well as the specific loan demand expression (2), describes the determination of variables that are unobservable in the presence of transactions costs. Consequently, it is necessary to apply some model of portfolio adjustment to translate the implications of such expressions into an operational model of behavior. Since transactions costs constitute, in the first instance, the underlying motivation for using a model which admits discrepancies between actual and desired portfolio holdings,[9] it is useful to distinguish the less costly (and hence more sensitive to asset yield and price expectations) allocation of the current investable cash flow ΔW_t from the reallocation of the investor's existing holdings W_{t-1}. An adjustment model that in the long run converges to the equilibrium given by (1), but in the short run incorporates this distinction in a tractable way, is the "optimal marginal adjustment" model[10]

$$(3)\quad \Delta A_{it} = \sum_{k}^{N} \theta_{ik}(\alpha_{kt}^{*} W_{t-1} - A_{k,t-1}) + \alpha_{it}^{*}\Delta W_t \qquad i=1,\dots,N$$

where the desired equilibrium proportions

$$(4)\quad \alpha_{it}^{*} \equiv \frac{A_{it}^{*}}{W_t} \qquad i=1,\dots,N$$

$$\sum_i \alpha_i^{*} = 1$$

follow from (1), and the θ_{ik} are fixed coefficients of adjustment satisfying $\Sigma_i \theta_{ik} = \bar{\theta}$ for all k, with $\bar{\theta}$ arbitrary.

Expanding portfolio adjustment model (3), using the desired loan demand expression (2) as a specific component of portfolio selection model (1), yields an operational equation for investors' short-run demand for loans:

$$\begin{aligned}(5)\quad \Delta L_t = {} & \pi_L \cdot \Delta W_t + \left[\sum_k (\pi_k \cdot \theta_{Lk})\right] \cdot W_{t-1} \\ & + \beta_{LL} \cdot r_{Lt}^e \cdot \Delta W_t + \left[\sum_k (\beta_{kL} \cdot \theta_{Lk})\right] \cdot r_{Lt}^e \cdot W_{t-1} \\ & + \sum_{k \neq L} \left\{ \beta_{Lk} \cdot r_{kt}^e \cdot \Delta W_t + \left[\sum_j (\beta_{jk} \cdot \theta_{Lj})\right] \cdot r_{kt}^e \cdot W_{t-1} \right\} \\ & + \gamma_{LL} \cdot v_{Lt} \cdot \Delta W_t + \left[\sum_k (\gamma_{kL} \cdot \theta_{Lk})\right] \cdot v_{Lt} \cdot W_{t-1} \\ & + \sum_{k \neq L} \left\{ \gamma_{Lk} \cdot v_{kt} \cdot \Delta W_t + \left[\sum_j (\gamma_{jk} \cdot \theta_{Lj})\right] \cdot v_{kt} \cdot W_{t-1} \right\} \\ & - \theta_{LL} \cdot L_{t-1} - \sum_{k \neq L} \theta_{Lk} \cdot A_{k,t-1}\end{aligned}$$

Here it is useful to distinguish the particular right-hand side terms which do and do not have coefficients of known sign a priori. Each expected holding-period yield r_k^e and variance $v_k, k=1,\dots,N$ (including the own yield r_L^e and own variance v_L) enters (5) twice, in non-linear form both times. In each case the product of r_k^e or v_k and the flow ΔW_t bears a coefficient which consists of a single parameter of known sign from (1). Similarly, the lagged own stock L_{t-1} enters (5) with coefficient $-\theta_{LL} < 0$ from the stock-adjustment component of (3). All other right-hand side terms in (5)—including the linear terms W_t, W_{t-1} and $A_{k,t-1}$, $k \neq L$, as well as all non-linear terms consisting of products of r_k^e or v_k with W_{t-1}—bear coefficients which are of unknown sign a priori.

Since the models of equilibrium portfolio allocation (1) and short-run portfolio adjustment (3) describe an investor's demands for

[9]See Duncan Foley for a useful analysis of this issue.

[10]See my 1977 paper for the development of the optimal marginal adjustment model.

all assets (and supplies of all liabilities), loan demand equation (5) is implicitly an element of a set of demand equations which satisfy the specified "adding-up" constraints. By contrast, the more limited focus of this paper is on the nature of investors' demands for loans—more specifically, long-term loans—and on the partial-equilibrium implications of these demands for the relationship between expected price inflation and nominal interest rates. While it is not necessary to use constrained estimation techniques to guarantee that the parameter estimates of the full set of demand equations satisfy the adding-up constraints, so that there is no inconsistency involved in estimating only one demand equation rather than the entire set, in principle a complete model including all investors (and securities issuers) and all assets (and liabilities) would be preferable in that it would not only facilitate a general-equilibrium analysis but also permit the researcher to examine the implications for other asset demand equations of the presence of a given variable in any one asset demand equation.[11] The construction of such a complete model, however, lies well beyond the scope of this paper.

Section II presents the results of estimating loan demand equation (5) applied to the demand for long-term fixed-interest loans by six separate categories of *U.S.* lenders.

II. Estimation Results

A. *Preliminary Issues*

It is useful at the outset to comment briefly on several aspects of the specification and estimation of the loan demand equations presented below.

Disaggregation: These equations represent the demand for long-term loans to private borrowers (that is, long-term corporate bonds) by six categories of *U.S.* lenders which together held approximately 94 percent of all such loans outstanding in the United States as of year-end 1978: life insurance companies (37.5 percent), other insurance companies (4.3 percent), private pension funds (12.7 percent), state and local government retirement funds (19.0 percent), mutual savings banks (5.2 percent), and households[12] (15.0 percent). This disaggregation is useful because such diverse lenders, facing different legislative and regulatory constraints and playing different roles in the markets' highly complex intermediation structure, are unlikely to exhibit identical portfolio responses to expected price inflation. For example, although life insurance companies must earn real dividends, since their liabilities are almost exclusively in nominal form it is not obvious that their asset demands are highly sensitive to expected real yields in the short run. In addition, life insurance companies in most states can invest only a small fraction of their portfolios in equities. By contrast, many pension funds face liabilities which are either explicitly or implicitly indexed to consumer goods prices, and private pension funds in particular have substantially more latitude in allocating their portfolios.

Data: The primary data source for the stock and flow quantities used to estimate these equations is the Federal Reserve System's flow-of-funds accounts.[13] These data are seasonally adjusted and are denominated in millions of dollars. The particular interest rate used for r_L is the observed nominal new-issue yield on long-term bonds issued by utility companies rated Aa by Moody's Investors Service, Inc. The Aa rated utility bonds provide the greatest continuity in terms of the frequency of new issues; they are also most representative of new-issue activity in the *U.S.* market. Previous studies of long-term nominal interest rate determination using the reduced-form

[11] William Brainard and James Tobin have argued forcefully along these lines.

[12] The household sector as defined here primarily consists of individuals but also includes nonprofit organizations and bank-managed personal trusts.

[13] See my 1977 paper for further details on precise definitions of variables, in particular the cash flows of life insurance companies and households, and dummy variables in the two insurance company equations.

term-structure approach have relied on indices of yields on either new issues or seasoned issues, but the new-issue yield is likely to be superior for several reasons including greater trading volume, fewer measurement problems, and absence of any term-coupon bias. The sample period consists of fifty-six quarterly observations beginning in 1960:I and ending in 1973:IV.

Instrumental variables estimation: Since the own yield on long-term loans is jointly determined by lenders' demands for loans and borrowers' supplies of loans, it is necessary to allow for this simultaneity in deriving consistent estimates of the loan demand equations. The relevant set of instruments used here for deriving consistent estimators includes not only the predetermined variables in the six respective disaggregated demand equations but also those in the two disaggregated loan supply equations developed in my 1979 paper. As is typically the case in multi-equation models, it is impossible to apply the two-stage least squares method directly because there are too many predetermined variables to permit ordinary least squares estimation of the system's reduced form as this method requires. The procedure used here follows James Brundy and Dale Jorgenson in using as instrumental variables not only the leading principal components of the full-system set of predetermined variables but also, on an equation-by-equation basis, the single equation sets of predetermined variables themselves.

Intercepts: Equation (5) has no intercept term, but it is probably accurate to consider not just (1) but (3) also as a linear approximation to a more complex behavioral pattern, and an intercept may follow from linearization. The procedure used here includes or excludes an intercept in each loan demand equation according to the standard error ratio.

Expectations proxies: Since lenders' expectations are unobservable, both for price inflation and for nominal holding-period returns on assets subject to capital gains or losses, it is necessary to use some indirect representation in their place. Nevertheless, no sharp consensus exists on the best form of expectations proxy to use for such purposes. The approach adopted here, therefore, is to estimate each of the six loan demand equations twice—once with an autoregressive expectations proxy used by many previous researchers, and once with a "rational" expectations proxy suggested by Bennett McCallum.

B. *Results Based on an Autoregressive Expectations Proxy*

A familiar representation of expectations, used by many researchers who have explored the relationship between expected price inflation and nominal interest rates, is that market participants form their expectations of relevant variables on the basis of previously observed values of these variables. Following Marc Nerlove, Jacob Mincer, and others, a general autoregressive expectation mechanism of the form

$$E_t(\tilde{x}_{t+1}) = \sum_{i=0} \omega_i x_{t-i} \tag{6}$$

where $E_t(\cdot)$ is the expectation held at time t, the ω_i are lag weights and the tilde indicates that variable $\tilde{x}_{t+1}$ is unknown as of time t, is consistent with the optimal linear prediction of a time-series from its own past history. Several familiar simple expectations mechanisms, such as the "naive" model that next period will be like last period, or Philip Cagan's adaptive expectation, are special cases of autoregressive expectations. More generally, Franco Modigliani and Robert Shiller have usefully illustrated, also in the context of interest rate and price expectations, that autoregressive expectations are consistent with a combination of extrapolative and regressive components. To emphasize the contrast to the more restrictive adaptive schemes, some writers have referred to such general autoregressive expectations mechanisms as "partly rational" or "weak-form rational."[14]

Table 1 shows the results of estimating loan demand equation (5), for each of the

[14]See, for example, Thomas Sargent, Rutledge (1974), Charles Nelson, and McCallum.

TABLE 1—LOAN DEMAND EQUATIONS BASED ON AN AUTOREGRESSIVE EXPECTATIONS PROXY

Life Insurance Companies

$$\begin{aligned}\Delta L_t^L = {} & \underset{(5.9)}{0.1716}\, W_{t-1}^L + \underset{(2.9)}{0.2193}\,(r_{Lt}\cdot\Delta W_t^L)^* - \underset{(-3.5)}{0.03203} \sum_i \phi_i \Delta r_{S,t-i}\cdot\Delta W_t^L - \underset{(-1.3)}{0.1215}\, r_{It}\cdot\Delta W_t^L - \underset{(-2.8)}{0.00216}\, r_{Ct}\cdot W_{t-1}^L \\ & - \underset{(-2.8)}{0.3225} \sum_i \zeta_i DP_{t-1}\cdot\Delta W_t^L + \underset{(2.9)}{0.00457} \sum_i \zeta_i DP_{t-i}\cdot W_{t-1}^L - \underset{(-1.7)}{2.241}\, v_{Lt}\cdot\Delta W_t^L \\ & + \underset{(1.4)}{0.02475}\, v_{Lt}\cdot W_{t-1}^L + \underset{(1.7)}{1.398}\, v_{Bt}\cdot\Delta W_t^L - \underset{(-5.6)}{0.3551}\, L_{t-1}^L - \underset{(-2.6)}{0.1405}\, I_{t-1}^L\end{aligned}$$

$\bar{R}^2=0.89$; $SE=154$; $D.W.=1.35$

Other Insurance Companies

$$\begin{aligned}\Delta L_t^O = {} & \underset{(2.1)}{524.1} + \underset{(2.8)}{0.2570}\, W_{t-1}^O + \underset{(6.2)}{0.0501}\,(r_{Lt}\cdot\Delta W_t^O)^* - \underset{(-1.4)}{0.00121} \sum_i \phi_i \Delta r_{S,t-i}\cdot\Delta W_t^O - \underset{(-7.0)}{0.02000} \sum_i \zeta_i DP_{t-i}\cdot W_{t-1}^O \\ & + \underset{(1.3)}{0.000109}\, v_E\cdot\Delta W_t^O - \underset{(-4.7)}{0.00000846}\, v_E\cdot W_{t-1}^O \\ & - \underset{(-2.4)}{0.2534}\, L_{t-1}^O - \underset{(-2.9)}{0.3005}\, M_{t-1}^O - \underset{(-2.8)}{0.2572}\, E_{t-1}^O - \underset{(-3.9)}{0.3985}\, U_{t-1}^O\end{aligned}$$

$\bar{R}^2=0.96$; $SE=42$; $D.W.=1.99$

Private Pension Funds

$$\begin{aligned}\Delta L_t^P = {} & \underset{(2.1)}{0.2892}\, W_{t-1}^P + \underset{(3.1)}{0.09953}\,(r_{Lt}\cdot\Delta W_t^P)^* - \underset{(-2.1)}{0.00445} \sum_i \phi_i \Delta r_{S,t-i}\cdot\Delta W_t^P - \underset{(-1.8)}{0.09811} \sum_i \zeta_i DP_{t-i}\cdot\Delta W_t^P \\ & - \underset{(-2.1)}{0.000428} \sum_i \psi_i DQ_{t-i}\cdot W_{t-1}^P + \underset{(3.0)}{1.323}\, v_{E't}\cdot\Delta W_t^P - \underset{(-3.2)}{0.02630}\, v_{E't}\cdot W_{t-1}^P - \underset{(-2.5)}{0.4633}\, L_{t-1}^P \\ & - \underset{(-2.0)}{0.2925}\, E_{t-1}^P + \underset{(3.7)}{0.3714}\, U_{t-1}^P\end{aligned}$$

$\bar{R}^2=0.63$; $SE=198$; $D.W.=2.48$

State-Local Retirement Funds

$$\begin{aligned}\Delta L_t^S = {} & -\underset{(-2.6)}{1834} + \underset{(3.4)}{0.2729}\, W_{t-1}^S + \underset{(2.0)}{0.07650}\,(r_{Lt}\cdot\Delta W_t^S)^* - \underset{(-2.8)}{0.01273} \sum_i \phi_i \Delta r_{S,t-i}\cdot\Delta W_t^S - \underset{(-1.4)}{0.04255}\, r_{Tt}\cdot\Delta W_t^S \\ & - \underset{(-2.4)}{0.1229} \sum_i \zeta_i DP_{t-i}\cdot\Delta W_t^S + \underset{(3.5)}{8.571}\, v_{Bt}\cdot\Delta W_t^S - \underset{(-3.8)}{0.2682}\, v_{Bt}\cdot W_{t-1}^S + \underset{(3.9)}{2.100}\, v_{E't}\cdot\Delta W_t^S \\ & - \underset{(-3.2)}{0.04604}\, v_{E't}\cdot W_{t-1}^S - \underset{(-3.0)}{1.376}\, v_{Pt}\cdot\Delta W_t^S - \underset{(-2.8)}{0.2674}\, L_{t-1}^S - \underset{(-3.0)}{0.1943}\, U_{t-1}^S - \underset{(-3.5)}{0.2511}\, E_{t-1}^S\end{aligned}$$

$\bar{R}^2=0.94$; $SE=95$; $D.W.=1.74$

Mutual Savings Banks

$$\begin{aligned}\Delta L_t^M = {} & \underset{(6.9)}{4906} + \underset{(6.0)}{0.3767}\,(r_{Lt}\cdot\Delta W_t^M)^* - \underset{(-4.1)}{0.02573} \sum_i \phi_i \Delta r_{S,t-i}\cdot\Delta W_t^M - \underset{(-4.4)}{0.4265}\, r_{Mt}\cdot\Delta W_t^M - \underset{(-1.5)}{0.07869} \sum_i \zeta_i DP_{t-i}\cdot\Delta W_t^M \\ & + \underset{(2.3)}{0.4147}\, v_{E't}\cdot\Delta W_t^M - \underset{(-3.7)}{0.01412}\, v_{E't}\cdot W_{t-1}^M - \underset{(-7.0)}{0.2722}\, L_{t-1}^M - \underset{(-6.1)}{0.3672}\, U_{t-1}^M + \underset{(6.9)}{0.5683}\, R_{t-1}^M - \underset{(-7.0)}{0.3317}\, C_{t-1}^M\end{aligned}$$

$\bar{R}^2=0.94$; $SE=97$; $D.W.=2.52$

Households

$$\begin{aligned}\Delta L_t^H = {} & \underset{(4.2)}{9995} + \underset{(5.1)}{0.06960}\,(r_{Lt}\cdot\Delta W_t^H)^* - \underset{(-2.2)}{0.00531} \sum_i \phi_i \Delta r_{S,t-1}\cdot\Delta W_t^H - \underset{(-4.0)}{0.09108}\, r_{Et}\cdot\Delta W_t^H + \underset{(4.4)}{0.00142}\, r_{Et}\cdot W_{t-1}^H \\ & - \underset{(-2.3)}{0.03282} \sum_i \zeta_i DP_{t-1}\cdot\Delta W_t^H + \underset{(4.1)}{0.00495} \sum_i \zeta_i DP_{t-i}\cdot W_{t-1}^H \\ & - \underset{(-2.2)}{0.00337}\, v_{Pt}\cdot\Delta W_t^H - \underset{(-4.0)}{0.2843}\, L_{t-1}^H - \underset{(-4.2)}{0.1533}\, U_{t-1}^H - \underset{(-4.2)}{0.00482}\, E_{t-1}^H\end{aligned}$$

$\bar{R}^2=0.86$; $SE=382$; $D.W.=2.94$

Notes: Numbers in parentheses are ratios of estimated values to standard errors. SE in millions of dollars.

Summary of Variable Symbols in Table 1

C = holdings of commercial mortgages
E = holdings of corporate equities
I = holdings of intermediate-term *U.S.* government securities
L = holdings of long-term corporate bonds
M = holdings of municipal bonds
R = holdings of residential mortgages
U = holdings of all *U.S.* government securities
W = total holdings of all financial assets
r_C = yield on new commercial mortgages (*ALIA* series)
r_E = yield on equities (Standard and Poor's dividend/price ratio)
r_I = yield on intermediate-term *U.S.* government securities (three–five years)
r_L = yield on corporate bonds (new Aa utility issues)
r_M = yield on municipal bonds (new Aaa issues)
r_S = yield on commercial paper (prime four–six months)
r_T = yield on Treasury Bills (three months)
v_B = variance of yield on long-term *U.S.* government securities
v_E = variance of yield on equities (dividend/price yield only)
$v_{E'}$ = variance of yield on equities (total return)
v_L = variance of yield on corporate bonds
v_P = variance of rate of price inflation (*CPI*)
DP = annualized percentage change of *CPI*
DQ = annualized percentage change of equity prices (Standard and Poor's)

six categories of lenders indicated above, using autoregressive proxies for the relevant expectations. The variable symbols are consistent for all six equations, with letter superscripts indicating distinctions among corresponding variables for the respective categories of lenders. Asterisk superscripts indicate terms for which an equation is estimated using fitted values of the variable from the first stage of the instrumental variables procedure. The numbers in parentheses are ratios of estimated coefficients to the corresponding standard errors.[15]

Following the autoregressive model, these equations use univariate distributed lags on past percentage changes of the *CPI*, past percentage changes in the one-period loan yield, and past percentage changes in equity prices to represent the influence of these past observations on lenders' expected holding-period yields. Similarly, these equations use computed moving-average variances to represent nonstationary second moments for the distributions describing lenders' expectations. The distributed lags on consumer prices and one-period loan yields are estimated, within the estimation of the loan demand equations, with lag weights constrained to follow a third-degree polynomial pattern, the right-hand tail of the lag constrained to pass through zero, and the lead lag weight free of the polynomial constraint.[16] The coefficients shown in Table 1 for the several distributed lag variables are in each case the sums of the estimated lag weights. The equity price distributed lags and the moving-average variances, by contrast, rely on uniform fixed weights.

The results shown in Table 1 are broadly consistent with the model of portfolio behavior developed in Section I. The estimated equations explain a large percentage of the variation of the changes of holdings (net purchases) of long-term loans by all lender categories other than private pension funds. The demand for loans in each case responds positively to the currently prevailing nominal yield on long-term fixed-interest loans; negatively to currently prevailing nominal yields on competing assets; negatively to the expected yield on a series of one-period loans, as represented by the distributed lag on past one-period loan yields; and negatively to expected price inflation, as represented by the distributed lag on past price movements. Private pension funds were the only lender group whose demand

[15]Because of the instrumental variables estimation procedure, the standard error ratios shown are asymptotically distributed as t-statistics but are not necessarily distributed as t-statistics in small samples.

[16]Nelson demonstrated that the joint estimation of the distributed lags together with the equations in which they are embedded (in contrast to estimating the lag structures independently and then including them in the corresponding equations) is the correct procedure. The estimation of these distributed lags presents an interesting identification problem due to the potential appearance of each distributed lag in two separate non-linear terms on the right-hand side of each single loan demand equation. See my 1977 paper with Roley for the derivation of the method used to solve the resulting estimation problem.

showed a significant response (as indicated by the standard error ratio) to the expected capital gain on equities represented by the fixed-weight distributed lag on past equity price movements; this term is excluded from the final specifications of the other five equations. The different moving-average variances enter these equations in an irregular way; but the own variance (calculated from realized net returns to holding corporate bonds) and the inflation variance enter negatively, while the variances of competing asset yields enter positively, whenever they appear in a product with the cash flow.[17]

Before proceeding to a closer inspection of these results, it is useful to consider potential inconsistencies introduced by the use of the autoregressive proxy for the unobservable expectations of price inflation. Charles Nelson has warned of the danger of inconsistent estimators in models that rely on autoregressive expectations proxies when actual expectations are not purely autoregressive, since in that case other independent variables in the regression may be correlated with the nonautoregressive portion of the actual expectation. In order to guard against this possibility, Nelson suggested independently constructing an autoregressive predictor for the time-series in question and calculating the correlations between the errors made by this predictor and past values of the model's other independent variables. To implement Nelson's test an autoregressive predictor of the rate of price inflation was constructed by ordinary least squares, using a twenty-quarter distributed lag with the lag weights constrained exactly as within the estimated loan demand equations. The correlations between the residuals for this independent price inflation equation and the past values of the independent variables shown in Table 1 (including both a one- and a two-period lag for any variable, like W_{t-1}, which is already lagged in the model) were significantly different from zero at the .05 level in only four cases out of ninety-three. This result for Nelson's test provides at least some empirical support for considering the autoregressive proxies potentially useful representations of lenders' unobservable inflation expectations.

Table 2 summarizes, for all six categories of lenders, three key coefficients of the estimated loan demand equation which are particularly relevant for assessing the role of lenders' portfolio behavior in the relationship between expected price inflation and nominal interest rates.

First, the estimated own yield coefficients $\hat{\beta}_{LL}$, from the specific loan demand equation (2) of the desired portfolio selection model (1), indicate the responsiveness of the desired fraction of loans in the portfolio. With the exception of mutual savings banks (for which $\hat{\beta}_{LL}$ is suspiciously large), these estimates are all of credible magnitude in addition to being significantly greater than zero at high confidence levels. In each case the value of $\hat{\beta}_{LL}$ indicates the fractional increase in the share of the portfolio that the lender will want to allocate to long-term fixed-interest loans if, with all other things equal, the nominal own yield on loans rises by 1 percentage point (that is, by 100 basis points).

Second, the estimated price expectations coefficients $\hat{\beta}_{LP}$, again from (2), indicate the responsiveness to expected inflation of the desired fraction of loans in the portfolio. These estimates are significantly different from zero, with the expected negative sign, for five of the six categories of lenders.[18]

[17]It is interesting that the standard error ratios of the variance terms are uniformly smaller (in absolute value) in these equations than in the corresponding ordinary least squares estimates. The results for the polynomial distributed lags are only modestly sensitive to the choice of lag length within the range of twelve to twenty quarters. See my 1977 paper and my forthcoming paper with Roley for further details of the exact procedures used in deriving the equations' final specifications.

[18]This coefficient was insignificant in the loan demand equation for nonlife insurance companies, and so the $(\Sigma\zeta_i DP_{t-i})\cdot\Delta W_t$ term is omitted from the final specification. That equation includes the price expectations distributed lag in the term $(\Sigma\zeta_i DP_{t-i})\cdot W_{t-i}$; but, as the discussion of (5) in Section I indicates, the coefficient of this term is a sum of products of parameters in the underlying model consisting of (1) and (3). The standard error ratio for the coefficient of the $(\Sigma_i\zeta_i DP_{t-i})\cdot\Delta W_t$ term in the equation for mutual savings banks is small, but the F-test indicates that this lag structure is significant at the .05 level.

TABLE 2—SELECTED PARAMETER ESTIMATES FOR LOAN DEMAND EQUATIONS BASED ON AN AUTOREGRESSIVE EXPECTATIONS PROXY

Lender Category	$\hat{\beta}_{LL}$	$\hat{\beta}_{LP}$	$\hat{\theta}_{LL}$	*RMSE*
Life Insurance Companies	.219 (2.9)	−.322 (−2.8)	.355 (5.6)	125
Other Insurance Companies	.050 (6.2)	—	.253 (2.4)	33
Private Pension Funds	.100 (3.1)	−.098 (−1.8)	.463 (2.5)	164
State-Local Retirement Funds	.077 (2.0)	−.123 (−2.4)	.267 (2.8)	77
Mutual Savings Bank	.377 (6.0)	−.079 (−1.5)	.272 (7.0)	84
Households	.070 (5.1)	−.033 (−2.3)	.284 (4.0)	323

Notes: Numbers in parentheses are ratios of estimated values to standard errors. *RMSE* in millions of dollars.

Hence the greater the expected inflation as inferred from recent observed inflation—all other things, including the nominal loan interest rate, equal—the smaller are these lenders' demands for fixed-interest loans. Following the discussion in Section I, this response may represent an explicit substitution of goods (or equities) for fixed-income assets, or it may indicate that investors draw inferences about future interest rate movements (hence about future capital gains) from observations of price inflation. The magnitudes of the $\hat{\beta}_{LP}$ estimates provide support for the hypothesis that lenders seek to maximize the utility of some real quantity since, for four of the six lender groups, it is impossible to reject at the .10 level the hypothesis $\hat{\beta}_{LP} = -\hat{\beta}_{LL}$. Nevertheless, since the estimated lag weight sums reported here are not identifiable as $\hat{\beta}_{LP}$ estimates without an arbitrary (though plausible) assumption that the true lag weights in the autoregressive expectation sum to unity,[19] it is not in general appropriate to assign these estimates a specific economic interpretation strictly comparable to the corresponding $\hat{\beta}_{LL}$. These estimates reflect the effect of observed price inflation on the demand for loans, including not only the effect of expected inflation on portfolio behavior but also the effect of observed inflation on expected inflation.[20]

Third, the $\hat{\theta}_{LL}$ estimates reflect the stock-adjustment component of the optimal marginal adjustment model (3). These estimates are significantly different from zero, with the expected positive sign, for all six categories of lenders. Their magnitudes roughly correspond to intuitive judgments of various lenders' respective likely speeds of portfolio adjustment based on institutional considerations; private pension funds, for example, which are typically managed very actively, undertake the most rapid

[19]The unit sum constraint implies that lenders believe that the stochastic process generating the price inflation is borderline stationary/nonstationary—that is, any rate of inflation which has persisted for a long time will continue to persist. The *U.S.* survey evidence suggests that inflation expectations in the 1970's do differ from those of the 1950's and 1960's in such a way as to render this borderline stationary/nonstationary specification plausible. Alternatively, for the process to be stationary, the lag weights in the estimated equations would have to sum to less than unity, and the expectation would also have to include a constant term.

[20]Another reason for caution in interpreting the $\hat{\beta}_{LP}$ estimates is the potential difficulty of distinguishing first- from second-moment effects. Myron Gordon and Paul Halpern, for example, have argued that the mean of the inflation rate is a good proxy for the associated uncertainty; such an effect here would bias upward the absolute values of the $\hat{\beta}_{LP}$ estimates. In addition, the estimated equations for the two categories of taxable lenders do not allow for specific tax effects, which have shifted during the sample period; see, for example, Martin Feldstein and Lawrence Summers.

reallocation of their existing assets.[21] Although these estimated adjustment speeds are fairly rapid in comparison to those typically found by previous researchers, they still indicate the existence of lags in portfolio behavior which will, in the short run, influence the response of nominal yields to any stimulus affecting lenders' behavior—including expected price inflation.

In addition, for purposes of comparison with the alternative set of loan demand equations presented below, Table 2 shows the root-mean square forecast error for each of the six equations.[22]

C. *Results Based on a Rational Expectations Proxy*

Recent researchers have broadly applied John Muth's concept of rational expectations. Expectations are rational, according to Muth's definition, if

$$\tilde{x}_{t+1} = E_t(\tilde{x}_{t+1}) + \tilde{u}_{t+1} \tag{7}$$

where $E_t(\cdot)$ is again the expectation conditional on all information available as of time t, and u is a zero mean finite-variance random disturbance which is serially uncorrelated as well as uncorrelated with $E_t(\cdot)$. In other words, expectations are rational in Muth's sense if the lender's expectation equals the mathematical expectation of the corresponding variable, conditional on all information available as of time t. As Jacob Frenkel and Michael Mussa among others have shown in the specific context of price inflation, if the nature of the process-generating realizations of the x series is such that all relevant information is contained in past values of x itself, then the autoregressive expectation as in (6) is also the rational expectation. Nevertheless, even in simple models the necessary conditions for rational and autoregressive expectations to be identical are typically severe. Hence it is useful also to estimate the loan demand equations using the expectations proxy suggested by McCallum explicitly for circumstances in which, because the generation of the variable(s) in question is more complicated than an autoregression, the autoregressive proxy is not a rational expectation.

In the absence of perfect foresight, Muth's definition of rationality renders the realization $\tilde{x}_{t+1}$ in (7) distributed around, rather than equal to, the expectation $E_t(\tilde{x}_{t+1})$, thereby leading to a classical errors-in-variables problem for estimation if actual values are simply used in place of the relevant expectations. Following McCallum, the procedure used here to estimate loan demand equation (5) under rational expectations therefore replaces the actual values corresponding to the three relevant expectations—of price inflation, capital gains on bonds, and capital gains on equities—with their respective Brundy-Jorgenson instrumented values based (either directly or through principal components) on all of the model's independent variables.[23] In the absence of any clear way to implement Muth's definition of rationality for higher moments

[21]This interpretation of the θ_{ii} estimates is merely heuristic, however, since in a multivariate stock-adjustment model the "speed of adjustment" depends on the eigenvalues of the entire matrix of θ_{ij} coefficients, not just the on-diagonal θ_{ii} values.

[22]For equations estimated by an instrumental variables procedure, the root-mean square forecast error (computed from the actual values of all right-hand side variables) is a better measure of statistical performance than is the estimated standard error (computed from instrumented values of the right-hand side variables).

[23]A check against the possibility that even the McCallum inflation expectations proxies might still fail to represent adequately Muth's rational expectations concept, because of the dependence of the true inflation process (and hence the rational inflation expectation) on yet further variables not included in the model's full set of independent variables, is to test the correlations between the implied expectations errors (that is, the proxy value minus the actual inflation rate) and past values of variables which might plausibly influence inflation but which are not among the model's independent variables. Correlations were therefore computed between the expectations errors implied by the McCallum inflation expectations proxy for each of the six loan demand equations and the lagged values of the growth rates of *GNP* in both current and constant dollars, the growth rates of the narrow (M_1) and broad (M_2) money stock, and the unemployment rate, as well as five successive lagged values of price inflation. None of the sixty correlations (ten lagged variables times one proxy for each of six loan demand equations) differed significantly from zero at the .05 level.

TABLE 3—SELECTED PARAMETER ESTIMATES FOR LOAN DEMAND EQUATIONS BASED ON A RATIONAL EXPECTATIONS PROXY

Lender Category	$\hat{\beta}_{LP}$	$\hat{\beta}_{LC}$	*RMSE*
Life Insurance Companies	−.0471 (−1.1)	.000437 (0.3)	221
Other Insurance Companies	.0517 (1.0)	−.00278 (−1.2)	129
Private Pension Funds	.0215 (1.2)	.0000375 (0.0)	190
State-Local Retirement Funds	.00740 (0.4)	−.00247 (−2.0)	98
Mutual Savings Bank	−.0369 (−3.6)	−.00339 (−3.3)	100
Households	−.0292 (−4.3)	−.0000555 (−0.1)	437

Notes: See Table 2.

of distributions, the variance representations included in these equations are the same moving-average variances used in the equations presented above based on the autoregressive expectations proxy.

Table 3 shows the estimates and standard error ratios for two coefficients of the loan demand equations estimated using the McCallum proxy. The estimated coefficients $\hat{\beta}_{LP}$ again indicate the responsiveness to the expectation of price inflation of the desired fraction of loans in the portfolio. The estimated coefficients $\hat{\beta}_{LC}$ analogously indicate the responsiveness of desired portfolio allocation to the expectation of capital gains on bonds. In sharp contrast to the corresponding estimates shown in Tables 1 and 2 for the loan demand equations based on the autoregressive proxy, these coefficient estimates are significantly different from zero, with the expected signs (negative for $\hat{\beta}_{LP}$, positive for $\hat{\beta}_{LC}$) for only two and zero categories of lenders, respectively.

Table 3 also shows the root-mean square forecast errors for each of the six loan demand equations estimated using the McCallum proxy. In each of the six cases, this error is greater than the error shown in Table 2 for the corresponding loan demand equation based on the autoregressive proxy. These root-mean square errors are essential to a valid comparison of the two sets of results, since high multicollinearity between the price inflation and capital gain series (as would be expected under perfect arbitrage) could in principle explain the weakness of the $\hat{\beta}_{LP}$ and $\hat{\beta}_{LC}$ estimates shown in Table 3. By contrast, the respective root-mean square errors are not sensitive to multicollinearity, and their comparison—which uniformly favors the equations based on the autoregressive proxy—is analogous to an F-test rather than the t-tests reported for the individual parameter estimates.

As a final test of the potential contribution of the McCallum proxy, the six loan demand equations were estimated using both the autoregressive proxy and the McCallum proxy for expected price inflation. The coefficient values and standard error ratios for the autoregressive proxies were typically smaller (in absolute value) than reported in Tables 1 and 2, but only marginally so. By contrast, only in the equation for life insurance companies was the coefficient of the McCallum proxy significantly different from zero with the expected negative sign. (In the equation for nonlife insurance companies the coefficient was significant but positive, and in the other four equations it was insignificant.)

Section III presents a partial-equilibrium analysis of the role of lenders' portfolio behavior in the relationship between expected price inflation and nominal interest rates, based on the loan demand equations estimated using the autoregressive proxy and shown in full in Table 1.

III. Expected Price Inflation and Nominal Interest Rates

The equations developed and estimated in Section II represent lenders' demand for loans. These six equations, together with some representation of borrowers' supply of loans, therefore constitute a complete partial-equilibrium model of the loan market. Adding the market-clearing equilibrium condition

$$\sum_{j=1}^{6} L_t^j = \bar{L}_t \tag{8}$$

where $\bar{L}$ is the supply of loans, thereby enables the model to determine the nominal loan yield which is an argument of each of the six estimated loan demand equations.[24] Furthermore, since five of these six equations explicitly include expected price inflation as another independent variable, the nominal loan yield determined in this model is an implicit function of expected inflation.

Figure 1 illustrates in (r_L, L) space how a partial-equilibrium analysis based on the six estimated loan demand equations and equilibrium condition (8), with loan supply taken as exogenous, isolates the contribution of lenders' portfolio behavior to the expected price inflation/nominal interest rate relationship. The object of this analysis is to show how the nominal loan yield would respond to expected inflation if all aspects of economic behavior other than lenders' portfolio behavior remained unchanged. In particular, the assumption of given investable cash flows (which are important arguments of the loan demand equations) holds unchanged all decisions about how much to save, and the assumption of a given loan supply analogously holds unchanged not only all decisions about how much to invest but also borrowers' decisions with respect to the composition of their liabilities.

[24]Since the six categories of lenders whose portfolio behavior is explicitly represented in the estimated loan demand equations do not hold all of the outstanding loans, $\bar{L}$ is more precisely the supply of loans minus those loans held by other lenders.

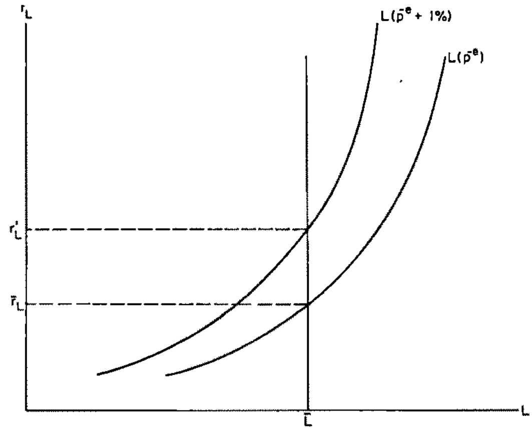


FIGURE 1. EFFECT OF LENDERS' PRICE EXPECTATIONS ON THE NOMINAL LOAN YIELD

In Figure 1 the upward-sloping curve $L(\bar{p}^e)$ represents the aggregated demand for loans conditional on some fixed expectation of price inflation $\bar{p}^e$. It intersects the fixed loan supply $\bar{L}$ at interest rate $\bar{r}_L$. Curve $L(\bar{p}^e+1\%)$, which represents the demand for loans conditional on an inflation expectation 1 percent greater, is shifted to the left from $L(\bar{p}^e)$, indicating lenders' reduced willingness to hold bonds at any given nominal interest rate. Curve $L(\bar{p}^e+1\%)$ therefore intersects $\bar{L}$ at interest rate $r'_L > \bar{r}_L$, and the difference $(r'_L - \bar{r}_L)$ (i.e., the "upward shift" measured by the vertical distance between $L(\bar{p}^e+1\%)$ and $L(\bar{p}^e)$ for a given $\bar{L}$) indicates the increase in r_L which makes lenders content to hold exactly $\bar{L}$ loans after an increase of 1 percent in their expectation of price inflation.

Figure 2 and Table 4 summarize the results of a dynamic version of such a partial-equilibrium analysis based on the six estimated loan demand functions from Section II, the market-equilibrium condition (8), and exogenous loan supply. The heavy solid line in Figure 2 plots the observed historical values of the nominal loan yield which, as Table 4 shows, averaged 6.07 percent over the 1960–73 sample period. The light solid line in the figure plots the simulated values of the nominal loan yield from the model of lenders' portfolio behavior. This "control" simulation, based on historical values of all

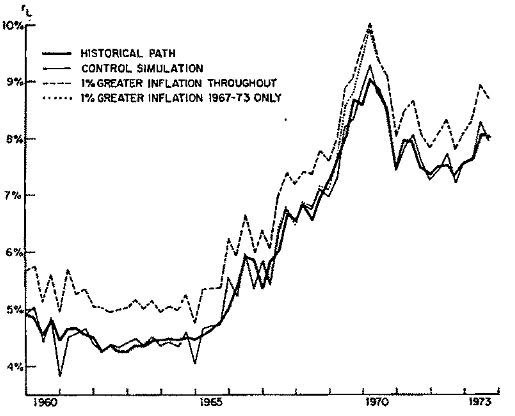


FIGURE 2. DYNAMIC SIMULATION RESULTS FOR THE LONG-TERM NOMINAL LOAN YIELD

exogenous variables (including loan supply, and distributed lags on observed price inflation in place of the unobservable inflation expectation), is fully dynamic in that, after the first quarter of the simulation, the solution uses internally generated values for the lagged own stock variables L_{t-1} in each of the six loan demand equations.

This control simulation indicates that the partial-equilibrium loan market model reproduces the relevant historical experience with reasonable accuracy. There is no significant bias for any of the model's seven jointly determined variables, and, as Table 4 shows, the mean simulated value of the nominal loan yield is precisely 6.07 percent. For the six loan demand variables, the root-mean square dynamic simulation errors are about in line with those shown in Table 2, indicating that the errors made by individual equations have no observable tendency to compound one another.[25] For the nominal loan yield, the root-mean square dynamic simulation error is 0.21 percent (i.e., 21 basis points)—about comparable to the "fit" achieved by previous researchers who have directly estimated reduced-form equations for long-term nominal interest rates.[26] This within-sample performance seems quite creditable, especially since the methodology of the structural model does not estimate an equation directly for this yield but, instead, implies an equation for the yield which is restricted by the underlying structural hypotheses about lenders' portfolio behavior.[27]

TABLE 4—NOMINAL LOAN YIELD DYNAMIC SIMULATIONS (SHOWN IN PERCENT)

	1960–73 Average Yield	Difference from Historical
Historical	6.07	–
Control Simulation	6.07	0.00
1 Percent Greater Inflation Experiment	6.72	0.65

What equilibrium adjustment in the nominal interest rate, equivalent to the difference between r_L' and $\bar{r}_L$ in Figure 1, will lenders' portfolio behavior induce in response to greater expectations of price inflation? The broken line in Figure 2 plots the simulated values of the nominal loan yield from an alternative simulation which differs from the control only in that the expected rate of price inflation is 1 percent greater throughout the simulation period. Specifically, this simulation experiment is based on

[25]The root-mean square error values for the six lender categories, in their order of appearance in Table 2, are 120, 36, 176, 87, 114, and 247, respectively.

[26]For example, Modigliani and Shiller's preferred equation had $SE=0.13$ percent for the less volatile Aaa yield over the sample period 1955:III–1971:II, but reestimating the Modigliani-Shiller equation using the Aa yield and the 1960:I–1973:IV sample period leads to an equation with $SE=0.22$ percent. Similarly, Feldstein and Otto Eckstein's preferred equation had $SE=0.09$ percent for the Aaa yield over the sample period 1954:I–1969:II, but reestimating the Feldstein-Eckstein equation using the Aa yield and the 1960:I–1973:IV sample period leads to an equation with $SE=0.29$ percent.

[27]This point is especially relevant to the presence of other long-term yields in the estimated loan demand equations for several categories of lenders. Including other long-term yields as independent variables in an unrestricted equation with the bond yield as dependent variable would presumably increase greatly such an equation's fit. In the context of the structural model, however, the contribution of other long-term yields is restricted to their role in influencing the net purchases variables. See my 1977 paper for a discussion of the structural modeling methodology as specifically applied to the determination of long-term interest rates.

values of the rate of change of consumer prices which are 1 percent greater than the corresponding historical values. The historical values are changed for as many years prior to 1960 as needed to be consistent with the distributed-lag price expectations terms in the estimated loan demand equations. In all other respects this simulation is identical to the control.

Since the demand for loans by five categories of lenders responds negatively to the greater expectations of price inflation,[28] the nominal loan yield must rise as in Figure 1 if total loan demand is still to equal the same historical total loan supply. As the broken line in Figure 2 shows, the market-clearing level of the nominal yield in this experiment is strictly greater than the control simulation level throughout the simulation period. The mean simulated value of the nominal yield in this experiment (r_L'), as Table 4 shows, is 6.72 percent—an increase of 0.65 percent above the control simulation mean ($\bar{r}_L$).[29]

Finally, what is the dynamic path by which lenders' portfolio behavior will bring about this 0.65 percent equilibrium adjustment? The dotted line in Figure 2 plots the simulated values of the nominal loan yield from a further experiment in which the 1 percent increase in the assumed rate of price inflation, in comparison with the control, is effective only in 1966:IV and thereafter. Hence this experiment's results for 1960–66 are identical to those of the control simulation. As of 1967:I, however, the simulated nominal loan yield begins to rise above the control path. By 1971:I (i.e., after four years) the adjustment to the new equilibrium path, which is identical to that of the first simulation experiment, is essentially complete.

These simple partial-equilibrium experiments cannot, of course, represent the complete nature of the expected price inflation/nominal interest rate relationship. In the first instance, the intent motivating their construction is not to model all of the underlying economic behavior but rather to isolate the role of lenders' portfolio behavior. Borrowers' portfolio behavior, for example, presumably corresponds to a downward-sloping loan supply curve instead of the vertical $\bar{L}$ in Figure 1, so that the equilibrium adjustment of the interest rate will be greater (less) than 0.65 percent if the "upward shift" of the supply curve is greater (less) than the 0.65 percent found here for the aggregated demand curve.[30] This analysis also abstracts from the influences of saving and investment behavior, as indicated at the outset. In addition, as the discussion of Section I notes, even the lenders' portfolio behavior modeled here applies only to one market—that for long-term fixed-interest loans—rather than to the complete set of all asset and liability markets. Nevertheless, these partial-equilibrium experiments are instructive in showing that lenders' portfolio behavior is an important part of the Fisher relationship, that this behavior alone is likely to yield a large (but less than one-for-one) equilibrium adjustment of nominal interest rates to expected price inflation, and that the dynamic path toward the equilibrium adjustment involves a substantial time lag.

IV. Summary of Conclusions

The adjustment of nominal interest rates to expected price inflation is a central issue

[28]Since the $(\Sigma_i \zeta_i DP_{t-i})\cdot W_{t-1}$ terms which appear in the loan demand equations for three categories of lenders reflect these lenders' behavior in other markets, the solution used the historical P values for these variables and introduced the 1 percent increase only in the $(\Sigma_i \zeta_i DP_{t-i})\cdot \Delta W_t$ terms. See again the discussion in Section I.

[29]An alternative to taking means over simulated values is simply to solve the model using the sample period means of the exogenous variables. This reverse strategy yields an estimate of 0.70 percent instead of 0.65. The difference is due to the model's nonlinearity.

[30]The analogous experiments reported in my 1978 paper, which includes an endogenous representation of borrowers' loan supply behavior, indicate that the loan supply curve actually "shifts upward" by slightly less than 0.65 percent for an additional 1 percent of expected price inflation. Hence the net upward adjustment of the interest rate due to the combination of lenders' and borrowers' behavior is slightly less than that indicated by the analysis of lenders' behavior alone, and the resulting quantity L is slightly below the historical $\bar{L}$.

for monetary theory and policy. The role of portfolio behavior is especially interesting in this context because, of the different kinds of economic behavior which may underlie the Fisher relationship, it is the most plausibly flexible in the short run. Even so, since substitution into real assets is not a practical portfolio alternative for many investors, it is not obvious a priori how important lenders' portfolio behavior is in this relationship.

The empirical results presented in this paper indicate that lenders' portfolio behavior does play an important role in the expected price inflation/nominal interest rate relationship.

First, at the single equation level, the results provide evidence that, with all other things equal, five of the six major categories of lenders in the *U.S.* long-term fixed-interest loan market reduce their demands for loans in response to an increase in expected inflation. Even life insurance companies, whose liabilities are almost entirely in nominal form, respond to price expectations in this way.

Second, at the multi-equation partial-equilibrium level, the results indicate that, with all other things equal, this response by lenders will raise the equilibrium nominal loan yield by 0.65 percent in response to a 1 percent increase in expected inflation. The results also indicate that this 0.65 percent adjustment requires approximately four years for completion.

REFERENCES

W. C. Brainard and J. Tobin, "Pitfalls in Financial Model-Building," *Amer. Econ. Rev. Proc.*, May 1968, *58*, 99–122.

J. M. Brundy and D. W. Jorgenson, "Efficient Estimation of Simultaneous Equations by Instrumental Variables," *Rev. Econ. Statist.*, Aug. 1971, *53*, 207–24.

P. Cagan, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago 1956.

F. De Leeuw, "A Model of Financial Behavior," in James Duesenberry et al., eds., *The Brookings Quarterly Econometric Model of the United States*, Chicago 1965.

E. F. Fama, "Multiperiod Consumption-Investment Decisions," *Amer. Econ. Rev.*, Mar. 1970, *60*, 163–74.

______, "Short-Term Interest Rates as Predictors of Inflation," *Amer. Econ. Rev.*, June 1975, *65*, 269–82.

M. Feldstein and O. Eckstein, "The Fundamental Determinants of the Interest Rate," *Rev. Econ. Statist.*, Nov. 1970, *52*, 363–75.

______ **and L. Summers,** "Inflation, Tax Rules, and the Long-Term Interest Rate," *Brookings Papers*, Washington 1978, *1*, 61–99.

Irving Fisher, *Appreciation and Interest*, New York 1896.

______, *The Rate of Interest*, New York 1907.

______, *The Theory of Interest*, New York 1930.

D. K. Foley, "On Two Specifications of Asset Equilibrium in Macroeconomic Models," *J. Polit. Econ.*, Apr. 1975, *83*, 303–24.

J. A. Frenkel, "Inflation and the Formation of Expectations," *J. Monet. Econ.*, Oct. 1975, *1*, 403–21.

B. M. Friedman, "Financial Flow Variables and the Short-Run Determination of Long-Term Interest Rates," *J. Polit. Econ.*, Aug. 1977, *85*, 661–89.

______, "Who Puts the Inflation Premium into Nominal Interest Rates?," *J. Finance*, June 1978, *33*, 833–45.

______, "Substitution and Expectation Effects on Long-Term Borrowing Behavior and Long-Term Interest Rates," *J. Money, Credit, Banking*, May 1979, *11*, 131–50.

______ **and V. V. Roley,** "Identifying Identical Distributed Lag Structures by the Use of Prior Sum Constraints," *Annals of Econ. Soc. Measure.*, Fall 1977, *6*, 429–44.

______ **and** ______, "A Note on the Derivation of Linear Homogeneous Asset Demand Functions," mimeo., Harvard Univ. 1979.

______ **and** ______, "Investors' Portfolio Behavior Under Alternative Models of Long-Term Interest Rate Expectations: Unitary, Rational or Autoregressive," *Econometrica*, forthcoming.

M. Friedman, "The Quantity Theory of

Money: A Restatement," in his *Studies in the Quantity Theory of Money*, Chicago 1956.

M. J. Gordon and P. J. Halpern, "Bond Share Yield Spreads under Uncertain Inflation," *Amer. Econ. Rev.*, Sept. 1976, *66*, 559–65.

B. T. McCallum, "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," *Econometrica*, Jan. 1976, *44*, 43–52.

R. C. Merton, "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case," *Rev. Econ. Statist.*, Aug. 1969, *51*, 247–57.

J. Mincer, "Models of Adaptive Forecasting," in his *Economic Forecasts and Expectations*, New York 1969.

F. Modigliani and R. J. Shiller, "Inflation, Rational Expectations, and the Term Structure of Interest Rates," *Economica*, Feb. 1973, *40*, 12–43.

M. Mussa, "Adaptive and Regressive Expectations in a Rational Model of the Inflationary Process," *J. Monet. Econ.*, Oct. 1975, *1*, 423–42.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica*, July 1961, *29*, 315–35.

C. R. Nelson, "Rational Expectations and the Estimation of Econometric Models," *Int. Econ. Rev.*, Oct. 1975, *16*, 555–61.

M. Nerlove, "Distributed Lags and Unobserved Components in Economic Time Series," in William Fellner et al., eds., *Ten Economic Studies in the Tradition of Irving Fisher*, New York 1967.

J. Rutledge, *A Monetarist Model of Inflationary Expectations*, Lexington 1974.

______, "Irving Fisher and Autoregressive Expectations," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 200–05.

P. A. Samuelson, "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *J. Polit. Econ.*, Dec. 1958, *66*, 467–82.

______, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Rev. Econ. Statist.*, Aug. 1969, *51*, 239–46.

T. J. Sargent, "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," *Brookings Papers*, Washington 1973, *2*, 429–72.

K. Wicksell, "The Influence of the Interest Rate on Prices," *Econ. J.*, June 1907, *17*, 213–20.

Forecasting the Market for New Ph.D. Economists

By W. L. HANSEN, H. B. NEWBURGER, F. J. SCHROEDER, D. C. STAPLETON, AND D. J. YOUNGDAY*

In this paper we present a forecasting model of the labor market for new Ph.D. economists. The most significant characteristic of our model is its explicit introduction of wages as short-run equilibrators of supply and demand in the Ph.D. market and as determinants of the long-run supply of persons entering and completing graduate study in economics. The predictions we derive are contrasted with predictions from a fixed-coefficients model in which there is no wage response.

I. Background

In 1972 the late Allan Cartter projected that the supply of new Ph.D.s available for teaching would grow steadily through the 1970's while demand for new teachers would remain roughly constant. Over the 1980–85 period he estimated an average annual surplus of 817 economics Ph.D.s, equal to 54 percent of the average annual output of doctorates. Interestingly, this projected surplus is almost equivalent to the entire actual output of Ph.D.s (877) in 1972. Based on his results Cartter suggested various possible adjustments, primarily supply responses, that might occur.

The labor market has proven to be softer than even Cartter predicted. The growth in the output of new Ph.D.s has slowed dramatically; in 1975 only 810 Ph.D.s were awarded, contrasted to his prediction of 1,055.[1] Meanwhile, real starting salaries declined by roughly 15 percent from 1970 to 1975. Obviously, substantial adjustments have already occurred, adjustments alluded to by Cartter but not an integral part of his model. Cartter's model precluded the possibility of adjustments; in particular, he assumed that a fixed proportion of college graduates would enter graduate school in economics and that a fixed proportion of new economics Ph.D.s would seek employment in academia.[2]

Our major objective is to incorporate several adjustment mechanisms in a forecasting model of the market for new Ph.D. economists. First, we specify the role of wages in bringing about market adjustments. Rather than forecasting supply and demand separately (as Cartter did), we allow changes in wages to mediate excess demands or supplies. Second, our model shows the interconnections between two sectors offering employment to new Ph.D.s; separate supply and demand functions in

*University of Wisconsin-Madison; Department of Housing and Urban Development; Board of Governors of the Federal Reserve System; University of British Columbia; and Montana State University, respectively. The research reported here was initiated by the members of a graduate public finance seminar at the University of Wisconsin. Earlier versions were presented at the *U.K.-U.S.* Conference on Teacher Markets at the University of Chicago in December 1976 and at the American Economic Association Meetings in December 1977. We are particularly indebted to the members of the seminar for their early contributions, to Burton Weisbrod for his continued interest and valuable suggestions, to various discussants and referees for their comments, and to the Sloan Foundation for financial support.

[1]Cartter is not the first to have erred in forecasting the market for economists. In 1962, Ewan Clague and Morton Levine projected the award of 561 new Ph.D.s in 1969–70. In a companion paper Francis Boddy cited a National Education Association study projecting a demand for 900 new college teachers of economics in 1969–70. In fact, 955 Ph.D.s were awarded in 1969–70, with non-Ph.D.s no doubt filling whatever gap in teaching positions may have existed. Although neither paper reported on possible salary changes, the salaries of economists, like those of other academics, did rise sharply in response to the tight labor market prevailing through most of the 1960's.

[2]Fred Balderston and Roy Radner explored the sensitivity of the results of a Cartter-type model for academics in general but did not incorporate market responses. In fairness to Cartter, it should be noted that in a subsequent work (1976) he discusses in a qualitative way some of the expected market responses in academic labor markets but does not incorporate them into his models or estimate their effects.

the academic and government sectors are specified. Third, we investigate how current labor market conditions affect graduate enrollments in economics and, consequently, the number of Ph.D.s produced in the future; again a market response mechanism is incorporated in the model.[3]

The market response approach is presented and discussed in Section II. A fixed-coefficients model, which excludes the market responses outlined above, is also presented. In Section III results of sample-period simulations of both models are presented, evaluated, and compared. Forecasts of the market model are presented in Section IV and are compared with forecasts from the fixed-coefficients model.

II. Model

The forecasting model has two parts: *the labor market* for new Ph.D.s in economics, including supply and demand equations for Ph.D.s in the academic and government sectors, and *the graduate enrollment sector,* including equations for the numbers of new and continuing graduate students and Ph.D.s granted.

Specification and estimation of the empirical model was necessarily limited by the availability of data as well as computational considerations. For most of the equations only seventeen observations were available. The equations presented below have been chosen from a much larger set than actually estimated. Our selection was based on 1) conformity of the coefficient estimates to our a priori notions as to sign, and 2) performance in simulations over the sample period. The latter was based on (a) how well each equation tracked, (b) how each equation affected the tracking ability of the model as a whole, and (c) how groups of equations affected the tracking ability of the model as a whole.[4] Heavy reliance was placed on mean square errors, but an element of judgment necessarily entered into our selections. This methodology reflects our purpose of constructing a useful forecasting model.[5]

The model captures two important adjustments to changing conditions in the market for new Ph.D. economists not included in fixed-coefficients models: adjustment in the total supply of new Ph.D. economists, and short-run adjustments in the proportions of new Ph.D.s going into the academic, government, and (implicitly) other sectors.

A. *Labor Market for New Ph.D. Economists*

Four equations determine the supply of and demand for new Ph.D. economists in the academic and government sectors.[6] An unsuccessful attempt was made to include a business sector.[7]

On the supply side, new Ph.D.s choose their sector of employment on the basis of salaries:

$$(1)\quad ln(PHDA^s/PHD) = \alpha_0^1 + \alpha_1^1 \ln WA + \alpha_2^1 \ln WG + \varepsilon^1$$

$$(2)\quad ln(PHDG^s/PHD) = \alpha_0^2 + \alpha_1^2 \ln WA + \alpha_2^2 \ln WG + \varepsilon^2$$

[3]This response mechanism is the main feature of the well-known cobweb models of markets for highly educated manpower developed by Richard Freeman, as in his study of the market for new Ph.D. physicists. Elton Hinshaw and Charles Scott and Scott have estimated a Freeman-type model of the academic market for new Ph.D. economists.

[4]In particular, an alternative specification of the model based on the theoretical structure set forth in Hansen et al. was rejected because this form did not track as well over the sample period. This alternative specification differed primarily in that its market sector did not use the stock-adjustment mechanism employed here.

[5]Thus, *t*-statistics presented below must not be interpreted as tests of hypotheses.

[6]In our data 40.1 percent of 1976 economics Ph.D. recipients planned employment at educational institutions, 10.9 percent in government, 7.3 percent in business or industry, and 2.2 percent in the private nonprofit sector. The remaining 40.1 percent took positions outside the United States, planned postdoctoral studies, did not know their postdoctoral status, or did not reply to the question.

[7]Estimation and simulation of a model including the business sector produced results quite similar to those

TABLE 1— LABOR MARKET EQUATIONS[a]
(Sample Period: 1959–75)

Equation	Dependent Variable	Constant	Independent Variables					R^2
			ln WA	*ln WG*				
Academic Supply	*ln(PHDA/PHD)*	2.90 (1.4)	.51 (1.9)	−.89 (2.6)				.29
Government Supply	*ln(PHDG/PHD)*	−4.71 (.84)	−1.13 (1.5)	1.33 (1.42)				.13
			PHDA(−1)	*WA*	*EU*	*GRAD*	*RD*	
Academic Demand	*PHDA*	65.54 (1.4)	.60 (3.8)	−.087 (1.5)	.37 (2.2)	.015 (.4)	.0006 (.4)	.89
Academic Stock[b]	*KPHDA*	8554.		−.23	.98	.04	.002	
			PHDG(−1)	*WG*	*FNDX*	*FDX*	*SLX*	
Government Demand	*PHDG*	6.60 (.6)	.73 (4.9)	−.015 (2.5)	.0008 (0.0)	−.90 (2.4)	2.04 (1.4)	.86
Government Stock[b]	*KPHDG*	1269.		−.06	.0032	−3.47	7.84	

[a]Absolute values of t-statistics in parentheses.
[b]Coefficients derived from the demand equations; see text.

where *ln* indicates natural logarithm, *PHD* is the number of new Ph.D. economists, *PHDA* and *PHDG* are the numbers taking positions in academia and government, respectively, with the superscripts indicating supply, *WA* is the median real salary for new assistant professors in economics, and *WG* is the starting salary for new Ph.D.s in the federal government.[8] The disturbances, ε^1 and ε^2, are assumed to be serially uncorrelated with constant variance and uncorrelated with all the exogenous variables in the model.[9]

Estimated coefficients and t-statistics for these equations are presented in lines 1 and 2 of Table 1.[10] In 1975, 810 persons were granted economics Ph.D.s and 325 of them took academic jobs. A \$1,000 real increase in academic salaries (from a mean of \$10,656 in 1975), holding government salaries constant, would induce 15 more new Ph.D.s to enter academic jobs, implying a point elasticity of about 0.5. A \$1,000 real increase in government salaries (from a mean of \$13,161 in 1975) would reduce the supply to academia by 23, implying a cross-salary elasticity near unity in absolute value. The elasticities in the government supply equation, though somewhat larger, imply smaller changes in quantities because fewer

reported here for graduate enrollments, the total number of Ph.D.s, and the trend of real wages for new economics Ph.D.s. The behavior of quantity and price in the business sector itself, however, was less reasonable in simulation. Omission of this sector therefore does not materially affect the results of this paper.

[8]All monetary variables are deflated to 1972 dollars by the implicit *GNP* price deflator.

[9]Similar assumptions are made on the disturbances throughout, although in some cases, such as the partial adjustment demand equations, these assumptions may be inappropriate. Because of nonlinearities involved and because the labor market equations are simultaneous, no attempt was made to account for this.

[10]These equations were estimated by instrumental variables. The instruments used were fitted values from regressions of *lnWA* and *lnWG* on *lnPHD, lnEU, lnGRAD, lnRD, lnFNDX, lnFDX,* and *lnSLX*: the latter six variables are the logarithms of variables appearing in the demand equations discussed below. We view the four supply and demand equations as a fully recursive, simultaneous block of the entire model (see Henri Theil, p. 461). Therefore we treat *PHD, EU,* and *GRAD* as exogenous variables in estimating these equations even though they are endogenous elsewhere. Because of the nonlinearities in the simultaneous system, two- or three-stage least squares estimation is inappropriate.

Ph.D.s entered the government sector (88 in 1975).

The specification of the demand equations for new Ph.D.s is derived from a linear stock-adjustment model. The desired number of Ph.D. economists in each sector is viewed as determined by factor prices and output levels. In any given year the stock of Ph.D.s carried over from the previous year satisfies most of a sector's demand. We assume that the desired stock of economics Ph.D.s, K^*, in a sector is a linear function of a vector of factor prices and output levels, $\underline{X}$:

$$K_t^* = \beta_0 + \underline{\beta}'\underline{X}_t \tag{3}$$

Adjustments in the desired number of Ph.D.s are made by hiring new Ph.D.s. At time t, the demand for new Ph.D.s, Q, is a partial adjustment of the difference between the desired stock and the stock carried over from the previous year, K_{t-1}; i.e.,

$$Q_t = \lambda(K_t^* - (1-\delta)K_{t-1}) \tag{4}$$

where δ is the rate of attrition due to deaths, retirements, and job mobility, and λ is the partial adjustment coefficient. A value for the attrition rate δ is based on estimates by Douglas Adkins.[11] Conditional on the correctness of this value (.02), the above equations imply:

$$Q_t = \alpha_0 + \alpha_1 Q_{t-1} + \underline{\alpha}'_2\underline{Z}_t \tag{5}$$

where
$$\alpha_0 = .02\lambda\beta_0$$
$$\alpha_1 = .98(1-\lambda)$$
$$\underline{\alpha}_2 = \lambda\underline{\beta}$$
$$\underline{Z}_t = \underline{X}_t - .98\underline{X}_{t-1}$$

(See Jack Johnston, pp. 300–03, for the transformation used to obtain these equations.)

Based on these equations, the academic and government demand equations in our model are

$$PHDA^d = \alpha_0^3 + \alpha_1^3 PHDA_{-1} + \alpha_2^3 \Delta WA + \alpha_3^3 \Delta EU + \alpha_4^3 \Delta GRAD + \alpha_5^3 \Delta RD + \varepsilon^3 \tag{6}$$

$$PHDG^d = \alpha_0^4 + \alpha_1^4 PHDG_{-1} + \alpha_2^4 \Delta WG + \alpha_3^4 \Delta FNDX + \alpha_4^4 \Delta FDX + \alpha_5^4 \Delta SLX + \varepsilon^4 \tag{7}$$

where ΔWA equals $(WA_t - .98\,WA_{t-1})$, ΔEU and $\Delta GRAD$ are analogous adjusted differences for undergraduate enrollment in colleges and universities and for graduate enrollments in economics, ΔRD is an adjusted difference for federal research and development grants in economics, and $\Delta FNDX$, ΔFDX, and ΔSLX are adjusted differences for federal nondefense, federal defense, and state and local government expenditures. The superscript d indicates demand. All nonsalary variables are intended as proxies for outputs of the two sectors; *GRAD* may represent an input as well, since graduate students contribute both to educating undergraduates and economic research.

The estimated coefficients appear in the third and fifth lines of Table 1.[12] Based on the estimates of the partial adjustment coefficients obtained from the coefficient on lagged quantity, the other coefficients can be solved for point estimates of the coefficients of the desired stock equation. These values are displayed in the fourth and sixth lines. The implied adjustment coefficient in the academic sector is .38. A \$1,000 real increase in academic salaries reduces the desired stock of Ph.D. economists in academia by 230 for 1975 values; this implies an elasticity of demand for new Ph.D.s of $-.22$. The coefficient on undergraduate enrollments implies a marginal stu-

[11]Adkins presents data on both the stock of all Ph.D. economists (in all sectors) and the flow of new Ph.D. economists. Regressing the current stock minus the current flow on the lagged stock yields a coefficient of .98; i.e., 98 percent of the previous year's stock was "carried over."

[12]Instruments used for the wage variables in estimating these equations were fitted values from regressions of the wages on all of the output variables and the total number of Ph.D.s. See fn. 10.

dent-teacher ratio of about 980.[13] One additional Ph.D. economist is required for every 25 additional graduate students in economics. Finally, a $1,000 real increase in federal *R&D* funds implies an increase in the desired stock of Ph.D. economists of about .002.

The implied adjustment coefficient for the government sector is .26. The desired stock of Ph.D. economists in government is reduced by 58 for each $1,000 real increase in salaries. At 1975 values, the implied elasticity of demand for new Ph.D. economists is −.47. Federal nondefense expenditures and state and local expenditures have positive influences on government demand, but the estimated coefficient for defense expenditures is inexplicably negative.

The final equations in the labor market section of the model are market-clearing identities:

$$PHDA^s = PHDA^d = PHDA \tag{8}$$

$$PHDG^s = PHDG^d = PHDG \tag{9}$$

B. *Graduate Economics Enrollments and New Ph.D.s*

The graduate enrollment sector includes three stochastic equations determining the number of entering graduate students in economics, the number of continuing graduate students, and the number of Ph.D.s awarded. An identity combines entering and continuing students to yield total graduate enrollments.

Economic theory suggests that the number of students beginning graduate work in economics is simultaneously determined by a "demand for places" from potential students who compare the expected rate of return in economics relative to other fields with the costs of going to graduate school, and by the "supply of places" provided by academic institutions. This approach is difficult to implement in practice. Developing the supply function and finding an appropriate price variable also proved difficult. The equation used for first-time graduate economics enrollments incorporates only demand-side determinants of enrollments; but it should not be interpreted as a demand function because it does not include a price (tuition) variable.

The first-time graduate economics enrollment equation is

$$\begin{aligned} \ln ENT = \alpha_0^7 &+ \alpha_1^7 \ln (W/WL)_{-1} \\ &+ \alpha_2^7 \ln STIP_{-1} + \alpha_3^7 SSGE \\ &+ \alpha_4^7 \ln BA_{-1} + \varepsilon^7 \end{aligned} \tag{10}$$

where *ENT* is full-time, first-year graduate students in economics, *W* is a weighted average of academic and government starting salaries for economics Ph.D.s (*WA* and *WG*), *WL* is a mean wage for intermediate level lawyers, *STIP* is the proportion of graduate students in all fields supported by federal fellowships and traineeships, *SSGE* is a selective service dummy variable, equal to unity in years for which graduate deferments were available for entering students and draft calls were high, and *BA* is the number of college graduates. The weighted average salary *W* is explicitly given by

$$W = (WA \cdot PHDA + WG \cdot PHDG) / (PHDA + PHDG) \tag{11}$$

The lawyer's wage *WL* is intended to represent the opportunity cost of becoming an economist; while other alternatives might have been used, it was necessary to choose a single wage because of the limited degrees of freedom. The variable *STIP* is included to capture the effects of increasing federal support for graduate education in the post-Sputnik period and declining support after the late 1960's. The variable *SSGE* is included because of the possibility that many students pursued graduate study to avoid the draft.[14] The variable BA_{-1} represents

[13] Of course, not all undergraduate students take economics courses all the time, so the ratio in classrooms is substantially smaller.

[14] Between 1960 and 1965, the number of persons drafted rose by about two-thirds. In 1966 inductions rose by another *230* percent. Deferments for entering graduate students ended June 30, 1968. Therefore this dummy is unity from 1965–66 to 1967–68.

the pool of potential graduate students in economics.

The estimated coefficients are given in the first line of Table 2. The elasticity of entrances with respect to the relative salary is .73. Federal support for graduate education also has a positive coefficient, consistent with the hypothesis that federal subsidies to students did increase graduate enrollments during the 1960's. The coefficient on *SSGE* implies that first-time enrollments were increased by about 23 percent through the combination of draft pressure and the availability of deferments—a very substantial effect. Finally, the coefficient on BA_{-1} implies that a doubling of the pool of potential entrants will increase first-time enrollments by about 78 percent.

The two other stochastic equations in the graduate enrollments sector determine the number of continuing students *CON* and the number of Ph.D.s granted.[15] The equation for *CON* is derived by assuming that *CON* is determined by a distributed lag on past entrants with geometrically declining weights:

$$CON = \beta_0 + \beta_1 \sum_{i=1}^{\infty} \lambda^i ENT_{-i}$$

A Koyck transformation (see Theil, p. 259) gives

$$\text{(12)} \quad CON = \alpha_0^9 + \alpha_1^9 CON_{-1} + \alpha_2^9 ENT_{-1} + \varepsilon^9$$

where
$$\alpha_0^9 = (1-\lambda)\beta_0$$
$$\alpha_1^9 = \lambda$$
$$\alpha_2^9 = \lambda\beta_1$$

Least squares estimates of the coefficients and implied values of the parameters of interest appear in the second and third lines of Table 2.[16] The estimates imply that about 75 percent of entering students leave by the beginning of the third year of studies.[17] The sum of entering and continuing graduate students gives total graduate economic enrollment:

$$\text{(13)} \quad GRAD = ENT + CON$$

The number of new Ph.D.s in economics is determined by the number of students entering three, four, and five, years previously, as well as by two draft dummies:[18]

$$\text{(14)} \quad PHD = \alpha_0^{11} + \alpha_1^{11} ENT_{-3} + \alpha_2^{11} ENT_{-4} + \alpha_3^{11} ENT_{-5} + \alpha_4^{11} SSGC + \alpha_5^{11} SSGEND + \varepsilon^{11}$$

The variable *SSGC* takes the value one during the years when continuing graduate status entitled students to deferments and draft pressures were high. A priori, the sign of the coefficient on this variable is expected to be negative as students delay completion. The variable *SSGEND* takes the value one in the year following the end of deferments for continuing students. It is expected that the end of deferments eliminated the incentive to delay completion and caused an exceptionally large number of Ph.D.s to be awarded in the following year.

Coefficient estimates are reported in the fourth line of Table 2. The sum of the coefficients on past entrances implies a completion rate of about 22 percent. Contrary to expectations, the coefficient of the deferment dummy indicates a positive effect on Ph.D.s awarded. The coefficient on *SSGEND* is positive, as expected.

C. *Summary of the Model*

Seven stochastic equations, (1)–(4), (7), (9), and (11), two market-clearing equations, (5) and (6), and two identities, (8) and (10), make up the market response model. The

[15]Clearly the decision to continue or complete graduate studies should also be affected by the costs and rewards of doing so. Our attempts to specify and estimate such a relationship were not fruitful.

[16]The value of *CON* for 1970 was inexplicably large. Since estimation of this equation using the 1970 value yielded questionable parameters, it was excluded from the estimation.

[17]A substantial number of these students may have terminated their studies upon receipt of the Master's degree. Nonetheless, 75 percent seems to be on the high side.

[18]A student granted the Ph.D. in his/her fourth year of study entered three years previously.

TABLE 2—GRADUATE SECTOR EQUATIONS[a]

Equation	Dependent Variable	Constant	Independent Variables					R^2		Sample Period
			$\ln(W/WL)(-1)$	$\ln STIP(-1)$	$SSGE$	$\ln BA(-1)$				
Graduate Entrances	$\ln ENT$	3.05	.73	.14	.23	.78		.98	$D.W.=1.8$	1959–75
		(3.6)	(1.6)	(3.1)	(5.3)	(9.3)				
			$CON(-1)$	$ENT(-1)$	$DUM70$	$DUM70(-1)$				
Graduate Continuations	CON	47.08	.35	.73	.830	−155.		.99	$h=-.31$	1955–75
		(.6)	(2.4)	(4.4)	(4.8)	(.7)				
	Implied Coefficients	β_0	λ	β_1	β_2					
		72.3	.35	2.10	830.					
			$ENT(-3)$	$ENT(-4)$	$ENT(-5)$	$SSGC$	$SSGEND$			
Ph.D.s Conferred	PHD	221.5	.045	.092	.081	106.1	53.5	.95	1.5	1961–75
		(5.4)	(.7)	(1.0)	(1.0)	(2.6)	(.8)			

[a]Absolute values of t-statistics in parentheses; $D.W.$ is the Durbin-Watson statistic; h is Durbin's h-statistic.

[library stamp]

TABLE 3—FIXED-COEFFICIENTS MODEL: ESTIMATES[a]
(Sample Period: 1960–75)

Equation	Dependent Variable	Constant	Independent Variables				R^2	
Academic Supply	*PHDA/PHD*	.498						
Government Supply	*PHDG/PHD*	.069						
Academic Demand	*PHDA*		*PHDA*(−1)	*EU*	*GRAD*	*RD*		
		−7.1	.86	.23	−.017	.002	.90	$h=-.87$
		(.2)	(9.0)	(2.0)	(.7)	(1.4)		
Government Demand	*PHDG*		*PHDG*(−1)	*FNDX*	*FDX*	*SLX*		
		−3.8	.87	2.6	−.70	1.32	.75	$h=-1.14$
		(.2)	(4.9)	(1.2)	(1.5)	(.7)		
	ln ENT		*ln STIP*(−1)	*SSGE*	*ln BA*(−1)			
Graduate Entrances		1.82	.19	.23	.89		.96	*D.W.*=1.9
		(5.3)	(5.3)	(4.8)	(16.9)			

[a]See Table 2.

nine endogenous variables are *PHDA, PHDG, PHD, WA, WG, W, ENT, CON, GRAD, PHD*. There are eleven exogenous variables: *EU, RD, FNDX, FDX, SLX, WL, STIP, SSGE, BA, SSGC*, and *SSGEND*.

D. *Fixed-Coefficients Model*

To illustrate the effects of the market adjustments in the forecasts, we estimate an alternative, fixed-coefficients model. This model differs from the market response model in two respects: market clearing in the labor market sector is no longer assumed (i.e., equations (5) and (6) are dropped) and the wage variables, *WA, WG, W*, and *WL*, are dropped from the equations in which they appear. Coefficient estimates for the five new behavioral equations appear in Table 3.

The supplies of new economics Ph.D.s to the academic and government sectors are simply constant proportions of all new Ph.D.s: .498 and .069, respectively. The coefficient estimates for the demand equations are similar to those of the market model. In the academic sector, the coefficient on graduate enrollments changes sign, and in the government sector the coefficient on nondefense expenditures increases. The proportion of variance explained by the demand equation in the academic sector is almost unaffected, but omission of the salary variable in the government sector reduces R^2 from .86 to .75.[19]

For the graduate enrollment sector in the market model, only the equation describing the number of entering students depends on wages. When the wage variable was omitted, none of the coefficients changed sign, and all but the constant term were comparable in magnitude. The proportion of variance explained dropped by less than 2 percent.

In summary, omission of the salary variables does not greatly affect the results of estimation; the fixed coefficients model seems to be a reasonable model in itself and suitable for comparison to the market response model.

III. Simulations

While the estimated coefficients in the behavioral equations of both the market response and fixed-coefficients models conform reasonably well to our a priori notions, this does not establish that these models are appropriate for forecasting. The evaluation

[19]One less observation was inadvertently used in estimation of the fixed-coefficients academic demand equation than in the market model. This accounts for the paradoxical increase in R^2 when the salary variable is omitted.

TABLE 4—MEAN SQUARE ERROR ANALYSIS
(Period: 1960–75)

Criterion	Endogenous Variables								
	Market Model Simulation								
	ENT	*CON*	*GRAD*	*PHD*	*PHDA*	*WA*	*PHDG*	*WG*	*W*
RMSE	117	142	219	41	33	312	7.4	482	296
RMSE/Mean	.048	.052	.042	.061	.100	.027	.155	.039	.024
MSE/Variance	.027	.023	.018	.046	.134	.068	.150	.024	.068
	Fixed-Coefficients Model Simulation								
	ENT	*CON*	*GRAD*	*PHD*	$PHDA^s$	$PHDA^d$	$PHDG^s$	$PHDG^d$	
RMSE	141	132	219	39.6	39.0	29.4	11.2	10.4	
RMSE/Mean	.058	.048	.042	.059	.118	.089	.235	.218	
MSE/Variance	.040	.020	.018	.043	.185	.105	.346	.298	

Note: *MSE* is mean square error; *RMSE* is root-mean square error.

in the previous section was on an equation-by-equation basis. For forecasting purposes the dynamic interaction of the entire system of equations is important. This section presents the results of simulating the two models over the period 1960–75. The ultimate test of the models is their future forecasting performance. It is unlikely, however, that their performance will be good if they cannot simulate well.

Three measures of fit based on a mean squared error (*MSE*) criterion are presented in Table 4 for each of the nine endogenous variables in the market response model. Root-mean square simulation errors (*RMSE*s) are presented in the first row.[20] Entries in the second row are the *RMSE*s divided by the mean values of endogenous variables in the sample period. For six of the nine endogenous variables, the *RMSE* is 6 percent or less of the variable mean. The largest value is 16 percent. Each entry in the third row is the *MSE* divided by the variance of the variable over the simulation period. Analogy to the error sum of squares divided by the total sum of squares of a regression suggests a loose interpretation of one minus R^2.[21] By this criterion the goodness of fit exceeds 90 percent for six of the nine variables and is never less than 81 percent.

Another important characteristic of the market model is its ability to reproduce turning points in simulation of the historical data. Six of the nine endogenous variables in the model rose almost continuously during the 1960's, peaked (each in a single year) between 1970 and 1972, and then declined through 1975. The values of these variables from the simulation reproduced each of these turning points correctly. The series on entering graduate students in economics displayed two peaks, and each of these was reproduced in the simulation, as was the local minimum between them.[22]

The *MSE* measures of fit for eight endogenous variables in the fixed-coefficients model are also presented in Table 4. Since the academic and government markets for new Ph.D.s are not cleared by salaries, differing supply and demand estimates are derived in the simulation (denoted in Table 4 by $PHDA^s$, $PHDA^d$, $PHDG^s$, $PHDG^d$). In the academic sector the extremes of excess supply in simulation were -40 in 1967 and 87 in 1975; in the government sector they were -23 in 1965 and 6 in 1960. In neither

[20]The $MSE=(1/T)\Sigma(\hat{Y}_t-Y_t)^2$, where T is the number of time periods in the simulation and $\hat{Y}_t, Y_t$ ($t=1,T$) are the predicted and actual values, respectively.

[21]The error sum of squares from simulation may exceed the variance, so that subtraction from unity may result in a negative value.

[22]The number of new Ph.D.s taking employment in government diplayed three peaks (in a generally rising trend), of which two were reproduced by the simulation. Salaries in the government sector rose and declined without displaying a clear turning point, and this pattern was reproduced in the simulation.

sector was excess supply or demand absurdly large. The *MSE* criteria are approximately the same for the market and fixed-coefficients models, with two exceptions: the market model reproduces with considerably greater accuracy both the actual flow of Ph.D.s into government and the number of entering students. The fixed-coefficient model does not perform quite as well as the market model in reproducing turning points. One turning point in the first-time enrollment series was missed and the 1972 peak in Ph.D.s entering academia was missed by both the supply and demand equations. However, all remaining turning points were reproduced.

In summary, while the simulation provides evidence that the dynamic structure of the market-response model is superior to that of the fixed-coefficients model, the evidence is not overwhelming. Both models duplicate the historical data well, suggesting that they can provide meaningful forecasts.

IV. Forecasting the Market for New Ph.D. Economists

Before forecasting the endogenous variables of our model it was necessary to assume values for the exogenous variables. Those which appear to be most important in determining the forecasts are undergraduate enrollments (*EU*) and bachelor's degrees conferred (*BA*). From 1960 to 1975 these variables increased by 222 and 256 percent, respectively. The projections show dramatically slower growth in enrollments from 1976 to 1981 (at a 1.1 percent average annual rate) and declines thereafter. By 1989 there is a slight decrease over 1976. The *BA* projections display a similar pattern.[23]

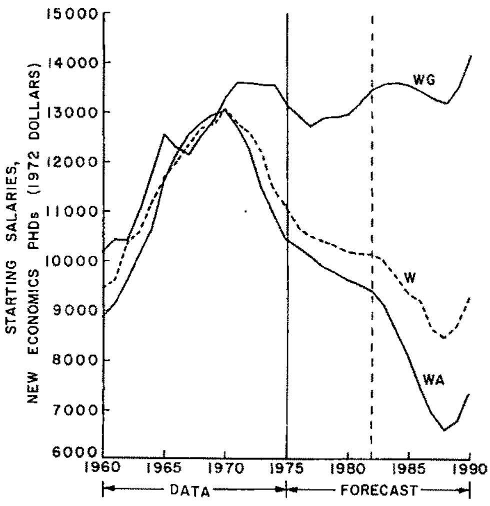


FIGURE 1

Historical values and forecasts of the endogenous variables are plotted in Figures 1 through 3. The solid vertical line at 1975 marks the beginning of the forecast period. After 1982 forecasted values for several variables are outside the range of variation found in the sample period and the broken vertical line indicates this. While we have less confidence in the post-1982 forecasts, we present them to show that the market model does not explode and that the dynamic character of the model produces reversals in some of the major trends in prices and quantities.

The results are striking. First, a pronounced decline occurs in real starting salaries of new economics Ph.D.s, over and above the already experienced 14 percent decline from 1970 to 1975. Real salaries drop by 17 percent from \$11,108 in 1975 to \$9,248 in 1990, hitting a low of \$8,483 in

[23]The U.S. Office of Education (*U.S.O.E.*) projects undergraduate enrollments *EU* and *BA* to 1985–86. We extended these projections to 1988–89 following as closely as possible the *U.S.O.E.* methodology. All the military draft dummy variables are set at zero. The percentage of graduate students receiving federal fellowships and traineeships (*STIP*) is held constant at the 1974 value (0.7), which is the historical low and follows a decline beginning in 1968. Federal research and development funds to economics (*RD*) grow at the same rate as real *GNP* and the salary variable for lawyers (*WL*) grows at the rate of *GNP* per capita. Forecasts for *GNP* as well as the three government expenditure variables (*FNDX*, *FDX*, *SLX*) come from a well-known macro-economic forecasting model. Average growth rates for all these variables for the past fifteen years and over the forecast period are

	RD	*WL*	*FNDX*	*FDX*	*SLX*	*EU*	*BA*
1960–75 (actual)	7.2	1.0	4.7	−1.0	4.8	6.2	6.4
1976–90 (assumed)	3.8	2.7	3.3	2.2	3.3	−0.3	−0.1

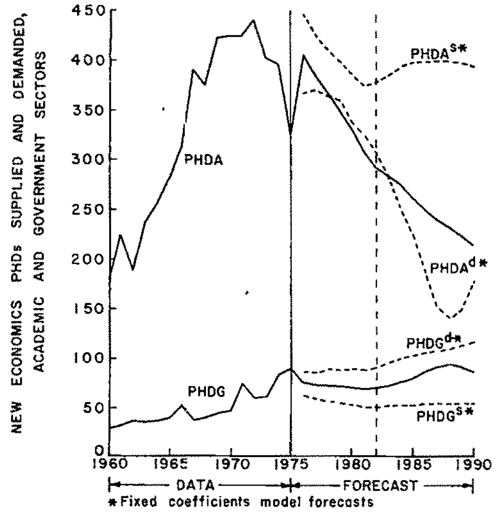


FIGURE 2

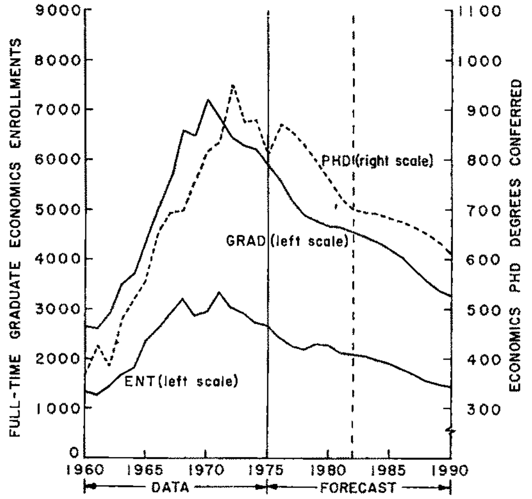


FIGURE 3

1988. Academic salaries fall more sharply, by about 30 percent, reflecting a reduced rate of growth and eventual decline in undergraduate enrollments and a drop in graduate economics enrollments. When undergraduate enrollments turn up in the late 1980's, starting academic salaries begin rising again. By contrast, real salaries in the government sector rise by about 7 percent over the forecast period, a reflection of their link to assumed favorable conditions for the economy as a whole.

Second, the number of new Ph.D.s drops sharply. Total new Ph.D.s decline steadily until the early 1980's, recover briefly, and then continue their decline through 1990. This decline is even more precipitous for Ph.D.s entering the academic sector; the forecast points to a steady drop from the 1975 level of 325. Meanwhile, employment in the government sector rises over the forecast period but at a slower rate than in the past.

The decline in new Ph.D.s reflects a monotonic and large decline in the number of persons entering graduate economics programs and, accordingly, a decline in total graduate economics enrollments. The decline in entering students occurs because of the decline up to 1984 in the average wage received by new economists in the academic and government sectors, and after 1984 because of the decline in *BA*s. These declines in graduate school entrances affect the supply side of the market with a four-year lag. Thus, the turnaround in academic salaries by 1989 occurs because the declining supply is just catching up with declining demand, and the decline in undergraduate enrollments is ending.

The sharp declines in salaries and new hires in the academic sector and the contrasting increases in salaries and new hires in government mark a dramatic change in the economics Ph.D. labor market. The percentage of new Ph.D.s obtaining jobs in government increases from 10.9 percent in 1975 to 13.7 in 1990. Moreover, the share of all Ph.D.s going into either academia or government declines from 53 percent in 1976 to 48 percent in 1990. This suggests a shift toward still other sectors, notably business, not included in the model.

A. *Assessing the Forecasts*

To assess the usefulness of the market response model, we compare the forecasts it produces with those of the fixed-coefficients model. While no rigorous criteria are available to judge the forecasts, several qualitative differences emerge that suggest the superiority of the market response model. We also

discuss some of the shortcomings of the market model and suggest how altering parts of it might change the results.

Fixed-coefficients forecasts of the main variables are depicted in Figure 3. Sustained excess supply of new Ph.D.s is predicted in the academic market and excess demand is predicted in the government market.[24] Annual excess supply in the academic sector is more than 50 percent of annual total supply from 1986 to 1990.

The market response forecast, in comparison, demonstrates clearly the effects of salary responses. Static salary adjustments (market clearing) lead to a large flow of new Ph.D.s into government—over 50 percent more by the late 1980's. Other Ph.D.s move into positions not explicitly modeled. Interestingly, lower salaries allow academic institutions to hire more new Ph.D.s, from 50 to 70 percent more in the late 1980's.

The most important difference between the two models is seen in the dynamic response of the total supply of new Ph.D.s to average starting salaries in the market model. Indeed, if the decision to enter graduate school, and thus the supply of new Ph.D.s, were not influenced by starting salaries, as in the fixed-coefficients model, real wages for new Ph.D.s would fall even lower than those predicted by the market response model.

While we conclude that the market response model is superior to the fixed-coefficients approach because its predictions differ significantly in a manner consistent with both a priori conceptions and actual observations of these labor markets, we recognize weaknesses that raise doubts about its forecasting accuracy. Many market adjustments are not explicitly included in our model. One such adjustment is increased hiring of new Ph.D.s by employers other than academic institutions and government. Another is possible changes in nonsalary job characteristics such as tenure policy[25] and research support.[26] Graduate school dropout rates are also likely to vary in response to market conditions.

Other weaknesses are also present. Different specifications of salary expectations could alter the speed of supply adjustments. Changes in the structure of the labor market over time may also affect the accuracy of our forecasts. The estimation period was dominated by growth in demand in the 1960's and a pattern of increasing inflation in the 1970's. While we forecast declining enrollments, the future of inflation is unclear. Moreover, the projected declines extend beyond the range of sample observations and lessen our confidence that the estimated equations are good approximations to the relevant structure beyond the early 1980's. Finally, the everwidening disparity in salaries between the academic and government sectors seems too great for us to accept with confidence.

Even without problems of misspecification or change in the underlying model, forecasts are always conditional on the assumed values of the exogenous variables which are, of course, themselves forecasts. The most important of these, undergraduate enrollments and bachelor's degree conferrals, come from a *U.S.O.E.* fixed-coefficients model—the very type of model we have criticized.[27]

V. Conclusion

We have presented a forecasting model of the market for new Ph.D. economists which incorporates two market-response mechanisms—in the short run, wages equilibrate supply and demand in two markets for new Ph.D.s, and in the long run,

[24]Both markets start out in disequilibrium in 1976 because of the fit of the fixed-coefficients model to the actual data in the early 1970's. Although adjusting the model so that the model is in equilibrium in 1975 would result in a forecast of near equilibrium in the late 1970's, there would still be substantial market imbalances in the 1980's.

[25]See Roy Radner and Charlotte Kuh.

[26]Real research support appears to be rising somewhat faster than our forecasts of it, as indicated by a 1977 National Science Foundation report.

[27]We did, however, try alternative forecasts for undergraduate enrollments based on Cartter's 1976 projections for these variables. The characteristics of the resulting forecasts were very similar.

salaries in these markets influence the number of graduate students and determine the aggregate supply of new Ph.D.s.

We conclude that real wages for academic economists will decline steadily until the end of the 1980's, when an upturn will occur; the number of new Ph.D.s taking academic jobs will decline steadily to 1990. Relative to the 1973–75 period, real salary declines will be on the order of 30 percent and new Ph.D.s getting academic jobs will drop about 35 percent. The situation in the government sector is less dismal, with real wages and the number of Ph.D.s finding jobs there rising slowly between 1973–75 and 1990. These changes reflect a decline in the share of new Ph.D.s going into academic jobs, an increase in the share going to government, and an increase in the share going to the private sector. We would anticipate some unemployment and underemployment to occur even though our model does not explicitly account for them. All of these events will lead to a contraction in the number of students entering economics programs. This will reduce the number of new economics Ph.D.s and cause the wage decline to be smaller than would otherwise be the case.

While we believe the market response forecasts are more realistic than those of the fixed-coefficients model, a true test is the relative accuracy of the predictions. For better or worse, any judgments about our model will have to be withheld for at least a few years. In the meantime some observers may believe our forecasts are too dire. In fact, the various market responses not modeled could entirely negate the forecasted changes.[28] If this should occur, additional responses will have to be incorporated into subsequent models. But if the prospects look dismal, and the projected values begin to materialize, then the economics profession, academic departments, and universities with graduate programs will have to decide whether to attempt to influence the market and produce an alternate outcome. What, if anything, should be done is left for readers to ponder.

[28]Of course, unforeseen changes in exogenous variables could invalidate the forecasts from either model. In this event, substitution of the correct exogenous variables and the development of new forecasts would provide a test of the models. Obviously, further work on forecasting the exogenous variables could be fruitful.

APPENDIX

Consistent time-series data were needed on market-clearing prices and the quantities of new Ph.D. economists entering different employment sectors. Boddy generously provided a special tabulation of median annual starting salary data for new assistant professors from his annual surveys of the major economics graduate degree-granting institutions for the period 1956–75. The National Research Council of the National Academy of Sciences furnished data on earned Ph.D. degrees and the employment plans of the degree recipients.

Variable Names and Data Sources

All monetary variables are deflated to 1972 dollars by the implicit *GNP* price deflator (see U.S. Council of Economic Advisors, *Economic Report of the President*).

BA: Bachelor's degrees conferred (in thousands). *Source*: U.S. Office of Education (*U.S. O.E.*), *Projections of Educational Statistics* (1977, Table 16, p. 33), and earlier editions.

CON: Continuing full-time graduate enrollment in economics. *Source*: *U.S.O.E.*, *Students Enrolled for Advanced Degrees*, various editions.

ENT: Entering (first-year) full-time graduate enrollment in economics. *Source: U.S.O.E., Students Enrolled for Advanced Degrees,* various editions.

EU: Full-time equivalent undergraduate enrollment in four-year institutions of higher education (in thousands). *Source:* calculated from total minus graduate enrollments

from *U.S.O.E., Projections of Educational Statistics* (1977, Tables 7 and 14, pp. 18, 25 for 1966–86; 1970, Tables 10 and 19, pp. 27, 36 for 1960–65; and 1967, Table 14, p. 21).

FDX: Federal defense purchases of goods and services (in billions of dollars). *Source: Economic Report of the President.*

FNDX: Federal nondefense purchases of goods and services (in billions of dollars). *Source: Economic Report of the President.*

GNP: *U.S.* gross national product (in billions of dollars). *Source: Economic Report of the President.*

GRAD: Full-time graduate enrollment in economics ($GRAD = ENT + CON$).

PHD: Doctor's degrees conferred in economics. *Source:* National Research Council, special tabulation.

PHDA;

PHDG: Current-year recipients of doctor's degree in economics taking employment in the academic and government sectors, respectively. *Source:* National Research Council, special tabulation.

RD: Federal research and development grants awarded in economics (in thousands of dollars). *Source: NSF, Federal Funds for Research and Development,* various editions.

SLX: State and local government purchases of goods and services (in billions of dollars). *Source: Economic Report of the President.*

SSGE; *SSGC*;

SSGEND: Dummy variables equal to unity for years in which at least 200,000 persons were conscripted and graduate students in economics were eligible for draft deferment (*SSGE* for entering and *SSGC* for continuing students); and a dummy variable equal to unity for the year 1969, the year after continuing student deferments ended (*SSGEND*). *Source:* U.S. Selective Service System, *Semi-Annual* and *Annual Report of the Director of Selective Service,* various editions.

STIP: Percentage of all graduate students supported by federal fellowships and traineeships. *Sources:* Freeman and David Breneman (p. 13) for graduate students supported and *U.S.O.E., Projections of Educational Statistics,* for total graduate students.

W: Weighted salary of new Ph.D.s in the academic and government sectors.

WA: Median salary of new assistant professors in economics. *Source:* special tabulation by Boddy.

WG: Starting salary of doctor's degree holders in federal civil service. *Source:* special tabulation by U.S. Civil Service Commission.

WL: Mean income of intermediate level lawyers (Level V from fiscal years 1960 to 1968, and Level IV thereafter to account for definition change). *Source:* U.S. Bureau of Labor Statistics, *National Survey of Professional, Administrative, Technical, and Clerical Pay,* various editions.

REFERENCES

Douglas L. Adkins, *The Great American Degree Machine*, Berkeley 1975.

F. E. Balderston and R. Radner, "Academic Demand for New Ph.D.'s: 1970–1990," in Roy Radner and Leonard S. Miller, eds., *Demand and Supply in U.S. Higher Education*, New York 1976, 315–49.

F. M. Boddy, "The Demand for Economists," *Amer. Econ. Rev. Proc.*, May 1962, *52*, 503–08.

Allan M. Cartter, "Whither the Market for Academic Economists," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 305–10.

______, *Ph.D.'s and the Academic Labor Market*, New York 1976.

E. Clague and M. Levine, "The Supply of Economists," *Amer. Econ. Rev. Proc.*, May 1962, *52*, 497–502.

R. B. Freeman, "Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973," *Amer. Econ. Rev. Proc.*, May 1975, *65*, 27–39.

______ **and David W. Breneman**, *Forecasting the PhD Labor Market*, Washington 1974.

W. L. Hansen et al., "The Market for Academic Economists: An Econometric Model," unpublished paper, Univ. Wisconsin, Nov. 1976.

C. E. Hinshaw and C. E. Scott, "The Market for PhDs (With Particular Reference to Economists)," Vanderbilt Univ., Sept. 1977.

Jack Johnston, *Econometric Methods*, 2d ed., New York 1972.

R. Radner and C. V. Kuh, "Market Conditions and Tenure in U.S. Higher Education," tech. rep. no. 2, Carnegie Council on Policy Studies in Higher Education, Apr. 1977.

C. E. Scott, "The Market for Ph.D. Economists: The Academic Sector," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 137–42.

Henri Theil, *Principles of Econometrics*, New York 1971.

National Research Council, *Summary Report 1976: Doctorate Recipients from United States Universities*, National Academy of Sciences, Washington, Mar. 1977.

National Science Foundation, (*NSF*) *Federal Funds for Research and Development*, Washington, various issues.

______, *Science Resources Studies Highlights*, Washington 1977.

U.S. Bureau of Labor Statistics, *National Survey of Professional Administrative, Technical, and Clerical Pay*, Washington, various issues.

U.S. Council of Economic Advisor, *Economic Report of the President*, Washington 1977.

U.S. Office of Education, (*U.S.O.E.*) *Projections of Educational Statistics*, Washington, various issues.

______, *Students Enrolled for Advanced Degrees*, Washington, various issues.

U.S. Selective Service System, *Semi-Annual and Annual Report of the Director of Selective Service*, Washington, various issues.

The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes

By STEVEN D. BRAITHWAIT*

The most commonly used measure of the cost of living, the Consumer Price Index (*CPI*) published by the Bureau of Labor Statistics, is essentially a Laspeyres price index. It is well-known, however, that a Laspeyres price index provides an upward biased estimate of the cost of living, because in keeping the same base period basket of goods as weights, it does not take into account substitution among commodities induced by relative price changes. The magnitude of the bias is an empirical question, which is the focus of this study.

An estimate of the substitution bias may be obtained by taking the difference between the Laspeyres price index and an estimated true cost-of-living index. The true cost-of-living index (*CLI*) is based on the theory of consumer demand. It is the ratio of the minimum expenditures under two different price regimes necessary to maintain a constant level of utility (as opposed to the constant basket of goods in the Laspeyres index).[1] The size of the bias depends on two factors: 1) the size of consumption substitution elasticities (if there is no commodity substitution then the *CLI* and Laspeyres index coincide, regardless of changes in relative prices); and 2) the magnitude of relative price changes (if all prices move together then the *CLI* and Laspeyres index again coincide, regardless of the size of substitution elasticities).

In order to compute a true *CLI* it is necessary to specify a particular utility function and estimate the parameters of the resulting system of demand equations (see, for example, Louis Phlips, p. 135). Several estimates of the magnitude of the substitution bias have appeared in the literature, but all previous studies suffer from a common weakness—they are all based on a very small number of commodities, essentially because of econometric problems involved in estimating large systems of demand equations. In one group of studies, demand models have been estimated using highly aggregated data, in which total consumption is grouped into four to nine aggregate "commodites" (such as housing, durables, or clothing). Arthur Goldberger and Theodore Gamaletsos, for example, estimate the linear expenditure system (*LES*) demand model using data on five aggregate goods for several European countries, and compare the resulting *CLI* for one of the countries (Greece) to the corresponding Laspeyres price index. Tran Van Hoa estimates cost-of-living indexes for six Australian provinces using data on nine aggregate goods. Phlips cites work by R. Sanz-Ferrer, who computed a *CLI* for Belgium from data on eight aggregate commodities.

In such studies using highly aggregated data, the estimate of the substitution bias may be understated because the aggregation obscures substitution in consumption within categories, which is perhaps more prevalent than substitution between gross aggregates. Another shortcoming of these three aggregate studies is the fact that they use only a single demand system, and hence do not explore the sensitivity of the *CLI* to a priori

*U.S. Bureau of Labor Statistics. The groundwork for this study was my dissertation. I would like to thank the chairman and members of my committee, R. Robert Russell, Llad Phillips, and Robert Deacon, for their support and helpful suggestions, and Jack Triplett, Robert Pollak, Richard J. McDonald, and a referee for valuable comments on an earlier version. The views expressed are my own and do not necessarily reflect the policies of the Bureau of Labor Statistics.

[1]For a theoretical discussion of cost-of-living indexes, see Robert Pollak (1971b), and Paul Samuelson and Subramanian Swamy.

specification of the form of the demand system.

The only previous studies which compare *CLI*s from several demand models (see Laurits Christensen and Marilyn Manser, and Manser) are, like the previously cited studies, carried out on a limited number of commodities. They are, however, based on relatively dissaggregated data and so are consistent with the approach of the present study. In one sense the Christensen-Manser results are more general than those reported here, since they employ a less restrictive functional form for some of the demand models. However, their results are limited by the use of data on food consumption only. Christensen-Manser compute cost-of-living subindexes for two commodity groups (meat and produce) of four commodities each. Manser presents similar results for a five-commodity breakdown of the food budget. Thus neither study obtains an estimate of an aggregate, all commodity *CLI*, or an estimate of the total substitution bias in the overall consumption price index.

The present study estimates the substitution bias in both the overall Laspeyres price index and in Laspeyres subindexes for various categories of commodities making up the consumption bundle. Cost-of-living indexes and subindexes constructed from the estimated parameters of three alternative consumer demand specifications are employed. In addition, I analyze the relative contributions of relative price variation and commodity substitution to the difference between the fixed-weight indexes and the true *CLI*s. Thus, this study extends previous work on the subject in that it combines the following factors: 1) the number of commodities is much larger than in any study previously reported in the literature, permitting the modeling of consumer behavior at levels of dissaggregation where substitution is more likely to take place; 2) in contrast with previous dissaggregated studies, the present one (nearly) exhausts the total consumption budget, allowing computation of the overall *CLI*, and the calculation of the bias in the price index for total consumption; 3) moreover, in addition to an overall *CLI*, cost-of-living subindexes, or category cost indexes, are calculated for each group of commodities into which the data are divided, thus providing an unique opportunity to examine the substitution bias of Laspeyres fixed-weight indexes for various parts of the consumer budget, as well as for total consumption; and 4) finally, cost-of-living indexes and subindexes are computed for three demand models, thus permitting evaluation of the sensitivity of the estimates to a priori specification of functional form.

The theory of the true cost-of-living index and subindexes needed to carry out the empirical work is presented in Section I. Section II presents the three alternative forms of cost-of-living indexes and subindexes which are estimated in this study. The data and empirical results are discussed in Section III. Section IV summarizes the results.

I. Cost-of-Living Indexes and Subindexes

This section summarizes some standard results from the theory of consumer demand and cost-of-living indexes, thus providing a theoretical framework for the empirical results in Section III. The exposition follows Robert Pollack (1971b).

A. *The True Cost-of-Living Index*

The most straightforward approach to the construction of a true cost-of-living index is through the consumer's indirect utility function.[2] Maximization of the more familiar direct utility function $U(X)$, where X is a list of commodities, subject to the budget constraint $P'X \leqslant Y$, yields a system of demand functions,

$$X=\phi(P,Y) \tag{1}$$

Substituting the demand functions (1) back into the direct utility function yields the indirect utility function,

$$V(P,Y)=U[\phi(P,Y)] \tag{2}$$

[2]For a discussion of the properties of indirect utility functions, see Lawrence Lau.

expressing maximum utility for a given set of prices and income (expenditure). Solving (2) for total expenditure Y yields the expenditure function $m(U,P)$, defined by

$$(3) \qquad Y = m(U,P)$$

which gives the minimum level of expenditure necessary to attain utility level U, given prices P.

Using equation (3), the cost-of-living index may now be defined as the ratio of the minimum expenditures under two different sets of prices necessary to maintain the level of utility in some base period:

$$(4) \qquad CLI(P_1,P_0) = \frac{m(U_b,P_1)}{m(U_b,P_0)}$$

Using Pollak's terminology, prices P_1 and P_0 are comparison prices and reference prices, respectively, and U_b is utility in the base period at prices P_b and income Y_b. The *CLI* is defined in reference to the utility function U, and in general will depend on the level of utility U_b in the base period. If the base period is chosen to be the same as the reference period, the denominator in (4) becomes actual expenditure Y_0, although base and reference periods need not coincide.

In contrast, the Laspeyres price index is defined as

(5)

$$L(P_1,P_0) = \Sigma P_1X_0/\Sigma P_0X_0 = \Sigma P_1X_0/Y_0$$

the ratio of the costs of the reference-period bundle of goods under the two different sets of prices. It maintains the base period level of consumption of each commodity, while the true *CLI* allows adjustment of the level of consumption of commodities in response to change in relative prices, while maintaining a constant level of utility. In fact, the Laspeyres price index provides an upper bound to the true cost-of-living index based on the reference period utility level (see Pollak, 1971b, for a proof).

B. *Category Cost Indexes*

In addition to indexes of the cost of living, it is of interest to examine the cost of subsets of commodities (i.e., the cost of clothing). This notion has a natural economic interpretation if the direct utility function is appropriately structured. Following Pollak (1975), if a subset X^s of the goods in the system is separable from all other goods X^c, then it is possible to define a "partial" cost-of-living index, or "category cost index," for that category of goods. Furthermore, if the utility function is weakly separable in the partition $[X^1,X^2,\dots,X^m]$ of the n goods, then it is possible to define an exhaustible set of category cost indexes corresponding to the groups $X^1, X^2,\dots,X^m$.[3]

In the weakly separable case the utility function may be written as

$$(6) \qquad U(X) = F[f^1(X^1),\dots,f^m(X^m)]$$

where f^s is the *category utility function* for the commodity subset X^s and F is the overall utility function, or aggregator function, combining the category utility functions. The category cost index (*CCI*) for the category of goods X^s may now be defined analogously to the *CLI*, using terms from the category utility function. Thus, V^s is the category indirect utility function defined analogously to (2) by

$$(7) \qquad V^s(P^s,y^s) = \max\{f^s(X^s)|\Sigma p^sx^s \leqslant y^s\}$$

where p^s and x^s are elements in the respective price and quantity vectors for the goods in the sth category, and y^s is category

[3]Pollak (1975) provides a thorough discussion of the restrictions on preferences which are necessary to define cost-of-living-type indexes for categories of commodities. He defines those subindexes which are independent of the level of consumption of goods outside the category (our category cost indexes) as partial cost-of-living indexes. W. Erwin Diewert and Charles Blackorby and R. Robert Russell discuss alternative types of cost-of-living subindexes derived from different assumptions about the structure of the utility function.

expenditure. Solving (7) for category expenditure yields the category expenditure function $m^s(f^s, P^s)$. The *CCI* is then the ratio of category expenditure functions:

$$(8) \qquad CCI^s(P_1^s, P_0^s) = \frac{m^s(f_b^s, P_1^s)}{m^s(f_b^s, P_0^s)}$$

In the present study the total list of commodities is partitioned into natural groupings, and a set of corresponding *CCI*s is calculated, as well as an overall *CLI*. It should be noted that four of the groups are themselves partitioned into subgroups, resulting in a three-level system of partitions.

II. Alternative Cost-of-Living Indexes

Computation of a cost-of-living index requires the specification of a particular functional form for the utility function (6) and, consequently, for the demand functions (1). All computed *CLI*s are estimates of "the" true cost-of-living index to the extent that the chosen utility function represents true preferences. Accordingly, one issue in evaluating the substitution bias in a fixed-weight price index is whether estimates of the bias are sensitive to the utility function chosen for the empirical work. To consider this issue, three different utility functions are employed in the this study, with *CLI*s and *CCI*s computed for each specification. The choice of utility functions in the present case is made by balancing realism, simplicity, and practicality in estimation. The three alternative sets of *CLI*s presented here correspond to multilevel versions of three well-known consumer demand models: the linear expenditure system (*LES*); generalized linear expenditure system (*GLES*); and indirect addilog.[4]

A possible weakness of these three models is that they all belong to the class of demand functions characterized by Pollak (1972) as exhibiting "generalized additive separability," which implies that interactions between demands for commodities are somewhat restricted. Generalized additive separability implies that the prices of "other goods" enter the demand function for each good in the form of an index of normalized prices. As a result, the independent influence on demand of prices and income is limited. For example, direct additivity (a property of the *LES* and *GLES* models) implies that the ratio of cross-price elasticities of two goods with respect to a third is independent of the third good, and is equal to the ratio of income elasticities for the two goods. In addition, inferior goods and complementary goods are ruled out.

More flexible models, such as the translog (see Christensen, Dale Jorgenson and Lau), relax some of these restrictions, but are more complex to work with, especially with the multilevel structure used here.

Studies using single level versions of the *LES*, *GLES*, and indirect addilog models are relatively familiar, but computational considerations greatly limit the number of goods that can be estimated with single level systems. To facilitate a more realistic modeling of consumption, the present study specifies weakly separable overall utility functions, resulting in multilevel versions of these three demands models, in which the levels correspond to the levels of the commodity partitions.[5] The overall utility function F in (6), which serves as an aggregator function for the category utility functions, $(f^1, \ldots, f^m)$, is assumed to be of the same functional form as the category utility functions.[6] The relevant equations for the category and overall utility functions, and

[4]The linear expenditure system was proposed by Lawrence Klein and Herman Rubin. It has since been applied by Richard Stone and many others. The generalized *LES* was analyzed theoretically in Pollak (1971a). The indirect addilog developed by Hendrik Houthakker has also been used in many applications. For a survey of the consumer demand literature, see Alan Brown and Angus Deaton, or Anton Barten.

[5]The first empirical implementation of such multilevel demand systems appears to be Murray Brown and Dale Heien's estimation of their *S*-branch system, a special case of the multilevel *GLES*.

[6]Because the indirect addilog is defined in terms of the indirect rather than the direct utility function, the

TABLE 1—UTILITY FUNCTIONS AND COST-OF-LIVING INDEX

	LES	*GLES*	Indirect Addilog
Overall Utility	$U=\prod_s (f^s-c_s)^{a_s}$	$U=\left\{\sum_s \alpha_s(f^s-c_s)^{\rho}\right\}^{1/\rho}$	$V=\sum_s \frac{\beta_s}{(1-\delta_s)}(g^s)^{1-\delta_s}$
Category Utility	$f^s=\prod_{i\in I_s}(x_i-\gamma_i)^{b_i}$	$f^s=\left\{\sum_{i\in I_s}\beta_i(x_i-\gamma_i)^{\rho_s}\right\}^{1/\rho_s}$	$g^s=\sum_{i\in I_s}\frac{b_i}{(1-\sigma_i)}z^{1-\sigma_i}, \quad z_i=\frac{P_i}{y^s}$
Category Cost Index	$CCI^s(P_1,P_0)=\dfrac{\Sigma\gamma_i p_{i1}+\frac{\pi_1^s}{\pi_b^s}(y_b^s-\Sigma\gamma_i p_{ib})}{\Sigma\gamma_i p_{i0}+\frac{\pi_0^s}{\pi_b^s}(y_b^s-\Sigma\gamma_i p_{ib})}$	Same as *LES*, with $\pi_t^s=(\Sigma b_i p_{it}^{1-\sigma_s})^{1/(1-\sigma_s)}$	$CCI^s(P_1,P_0)=\frac{y_1^{*s}(g_b^s)}{y_0^{*s}(g_b^s)}$ where $y_t^{*s}(g_b^s)$ is the solution in y^s of implicit equation:
	$\pi_t^s=\prod_{i\in I_s}(p_{it}/b_i)^{b_i}$	$\sigma_s=\frac{1}{1-\rho_s}$, and $b_i=\beta_i^{\sigma_s}$	$g_b^s=\Sigma\frac{b_i}{(1-\sigma_i)}(p_{it}/y^s)^{1-\sigma_i}$
Cost-of-Living Index	$CLI(P_1,P_0)=\dfrac{\Sigma(c_k\pi_1^k+\Sigma\gamma_i p_{i1})+\frac{\Pi_1}{\Pi_b}\Sigma(m_b^k-c_k\pi_b^k)}{\Sigma(c_k\pi_0^k+\Sigma\gamma_i p_{i0})+\frac{\Pi_0}{\Pi_b}\Sigma(m_b^k-c_k\pi_b^k)}$	Same as *LES*, with $\Pi_t=\left\{\sum_k a_k(\pi_t^k)^{1-\sigma}\right\}^{1/(1-\sigma)}$	See text
	$\Pi_t=\prod_k(\pi_t^k/a_k)^{a_k}$, and	$\sigma=\frac{1}{1-\rho}$, and $a_s=\alpha_s^{\sigma}$	
	$m_t^s=y_t^s-\sum_{i\in I_s}\gamma_i p_{it}$		

formulas for the *CCI* and *CLI* are presented for all three models in Table 1.

The method of estimation for the three demand systems is described in detail in my related paper. Briefly, the multilevel demand systems are estimated in stages, an econometric procedure rationalized by what has come to be known as a "two-stage budgeting procedure," the validity of which derives from the assumption of weak separability of the utility function.[7]

III. Empirical Results

A. *Data*

The quantity and price data used in this study are based on annual time-series data from 1948 to 1973 on Personal Consumption Expenditures in the United States, published annually by the U.S. Department of Commerce. Quantities purchased are represented by expenditures in constant (1958) dollars, and prices are implicit price deflators (deflators were chosen over *CPI* price index components to maintain consistency with the quantity data). The data are deflated by population to yield per capita expenditures.

The complete system consists of fifty-three commodities divided into six main groups, four of which are further divided into ten subgroups, as shown in the Appendix. The grouping, though arbitrary, is based on a judgment of similarity of the goods in nature or use, and results in category cost indexes for natural and familiar groups of goods. However, no tests for separability are made. Expenditure on these commodities amounts to nearly 90 percent of total Personal Consumption Expen-

procedure for building up its multilevel version differs somewhat from that implied by equation (6). Weak separability of an indirect utility function does not imply weak separability of the corresponding direct utility function which is needed to validate the two-stage budgeting procedure and the calculation of *CCI*s. Thus the multilevel indirect addilog used here is defined by a weakly separable *conditional* indirect utility function, which is a composition of category indirect utility functions by a weakly separable overall function (all of the indirect addilog functional form in the present case). This construction assures that the corresponding direct utility function. though unknown, is also weakly separable. For a more detailed discussion of conditional utility functions see Blackorby, Daniel Primont, and Russell.

[7]The procedure followed rests on the thorough discussion of the interrelationships between functional structure, consumer budgeting and demand theory in Blackorby, Primont, and Russell.

diture (*PCE*), as defined in the National Accounts.[8]

B. *Cost-of-Living Indexes and Subindexes*

Cost-of-living indexes, and category cost indexes for the ten commodity subgroups and six main groups used in the study were computed for all three demands models[9] and for all years 1948–73, using several different base years (1948,1958, and 1973). The estimated *CLI*s and *CCI*s for the alternative models were found to be extremely similar, differing for the most part by no more than 1 percent over the fifteen-year period from 1958–73. Exceptions are three subgroups (Personal Business, Recreational Goods, and Entertainment) for which the indirect addilog indexes differ from those of the *LES* and *GLES* by 8–13 percent (although the latter two agree closely). The special properties of these three subgroups are considered below.

The similiarity of the alternative cost indexes is an indication of robustness of the *CLI*s to the choice of demand system, although such a conclusion is tempered by the restrictiveness of the three models, discussed earlier. On the other hand, the results of the present study are consistent with findings of Christensen-Manser and Manser, who reported close agreement among the values of the *CLI*s from all of their models. Both studies included the flexible translog model (which does not belong to the generalized additively separable class) as well as the *LES* and indirect addilog. It is not computationally feasible to estimate the translog model for the full fifty-three-commodity system employed in the present study; but considered together, the present study and the Christensen-Manser and Manser food studies suggest that estimates of true cost-of-living indexes are robust to choice of functional form.

Since the primary focus of this paper is on the differences between Laspeyres price indexes and cost-of-living indexes for a large number of commodity groups, and because of the similiarity of the results between models, the estimated *CCI*s and *CLI*s for only one model—the *LES*—are presented here. The *LES* model is much less complex than the indirect addilog in the multilevel form used in this study. Furthermore, tests of the generality of the *GLES* over the *LES*, which are reported in my related paper indicate that the *GLES* is not a significant improvement over the *LES* in more than half of the categories. Space dictates summarization of the main results for the *LES* (in Table 2), however time-series of all indexes for each model are available on request.

The first column of Table 2 presents *LES* cost-of-living and category cost indexes for 1973, with both base and reference year 1958. Thus, the table gives the percentage change in the indexes over the fifteen-year period from 1958 to 1973, slightly longer than the normal length of time between major weight revisions in the *CPI*. The second column presents an overall Laspeyres price index and Laspeyres subindexes for the groups and subgroups, for the same

[8]The data were obtained from the U.S. Department of Commerce's Bureau of Economic Analysis on computer cards and thus in some cases are at a finer level of detail than the published data. Some commodities were omitted completely. The interpretation of a few of the *PCE* items in terms of actual consumption transactions is difficult because the method of construction of the expenditures is influenced by bookkeeping conventions of the National Accounts (see Hendrik Houthakker and Lester Taylor, pp. 56-57). Nearly half of the omitted expenditures were for two such items—"expense of handling life insurance" and "services furnished without payment by financial intermediaries other than life insurance companies." Also omitted were expenditures on hospitals, private education, and religion and welfare, which actually represent the output of organizations in those categories rather than purchases from them.

[9]It was not possible to calculate some of the aggregate indexes for the indirect addilog system, owing to computational complexity. The *CCI* for the lowest level of aggregation (subgroups) for the indirect addilog can only be defined implicitly (see Table 1), and its calculation requires the solution of an implicit non-linear equation. The *CCI* for a group and the overall *CLI* are defined respectively by systems of m and n implicit non-linear equations where m is the number of subgroups in the group and n is the total number of subgroups and groups in the system (sixteen in the present case). The iterative routine used to solve these non-linear equation systems failed to converge.

TABLE 2—SUBSTITUTION BIASES AND 1973 VALUES OF COST-OF-LIVING INDEX, CATEGORY COST INDEXES, AND LASPEYRES PRICE INDEXES
(Base Year Utility (Weights for Laspeyres) = 1958)
(1958 = 100.00)

	CLI and *CCI*s for *LES*	Laspeyres Price Indexes	Substitution Bias
Total Consumption	146.0	147.5	1.5
Categories			
I. Food	155.4	156.1	.7
II. Clothing	148.5	149.3	.8
III. Shelter	135.8	136.8	1.0
1. Housing and Utilities	139.8	139.9	0.1
2. Household Furnishings	133.0	133.2	0.2
3. Household Durables	122.8	126.7	3.9
IV. Transportation	132.5	134.0	1.5
1. Auto Services	129.3	129.5	0.2
2. Public Transportation	178.4	181.2	2.8
V. Personal Service	168.1	172.6	4.5
1. Medical Care	164.4	161.7	−2.7
2. Personal Business	181.9	197.7	15.8
3. Personal Care	169.8	169.8	0.0
VI. Recreation and Entertainment	139.9	146.9	7.0
1. Recreational Goods	124.5	126.4	1.9
2. Entertainment	185.8	193.2	7.4

period.[10] The major interest of this study focuses on the bias in the fixed-weight Laspeyres price index, which can be derived simply by comparing the *CLI* (or *CCI*s as appropriate) with the appropriate Laspeyres index. The values of the estimated biases (the difference between the first and second columns) are presented in the third column of Table 2.[11]

Looking first at the indexes for total consumption, the estimated bias in the overall Laspeyres index over the fifteen-year period from 1958 to 1973 is 1.5 percentage points, or approximately one-tenth of one point per year. The estimates of substitution bias in the category price indexes show a considerable range; more than half of the biases amount to under two percentage points over the fifteen years, but for some groups they are considerably larger. Perhaps the major weakness of the National Accounts data used in this study is the fact that they do not permit detailed disaggregation of food consumption. The Christensen-Manser studies are thus complementary to the present study for they compute food *CCI*s for nearly the same time period as the indexes in Table 2. They report biases in their food categories which are comparable in size to the biases in the present study for overall consumption and for the "low bias" groups—around 0.1 points per year.

[10]The theoretical upper bounds on the cost indexes for a given demand model are Laspeyres price indexes using the model's *predicted quantities* in 1958 as weights. This is so because the use of a particular model to compute a *CLI* is predicated on the assumption that it is the true model. The Laspeyres indexes in Table 2 were thus computed using *LES* predicted quantities as weights.

[11]For subgroups V.1 and V.3, second-order conditions for utility maximization were violated by the parameters of all three models in all or part of the sample period, implying that the *CCI*s for those two subgroups cannot be considered true cost indexes. Thus, in these two cases, the theoretical conditions guaranteeing a nonnegative bias are not satisfied.

For almost any conceivable use of the indexes, biases on the order of 0.1 percentage points per year must be considered quite low. Rigorous methods to test for statistical significance of bias estimates, however, do not exist (see Anthony Lawrence for a statement of the problem).

It is of interest to examine the high-bias categories. The five subgroups with the largest biases are: Household Durables (3.9 index points); Public Transportation (2.8 points); Personal Business (15.8 points); Recreational Goods (1.9 points); and Entertainment (7.4 points). Three of the five high-bias cases thus amount to between two and four index points over the total fifteen-year period, or approximately 0.2 points per year. Personal Business and Entertainment have higher bias estimates.[12]

C. *Relative Importance of Substitution and Changes in Relative Prices*

It seems natural to expect that the greater the degree of substitution and the greater the amount of relative price change within a category, the greater will be the bias in a fixed-weight Laspeyres index. As already noted above, if all prices change proportionately (i.e., no relative price change) the true *CLI* and the fixed-weight Laspeyres index change by identical amounts, and there is zero bias. Likewise, the bias is zero if there is no commodity substitution, even if there are large changes in relative prices. The Laspeyres index is the *CLI* for this case (see Pollak 1971b). In general, the bias is an interactive function of relative price changes and commodity substitution. It is accordingly of interest to determine whether the high-bias cases in Table 2 are associated primarily with large relative price changes or with high degrees of commodity substitution.

A detailed investigation of these two factors would require the examination of numerous statistics for each category, including measures of dispersion for each price relative, and a matrix of Allen partial elasticities of substitution, or income-compensated price elasticities. However, due to the large number of biases to examine, more compact measures of these factors are desirable.

A concise measure of the degree of commodity substitution is the expenditure share-weighted average of income-compensated own-price elasticities. (Income-compensated price elasticities are related to the Allen elasticities of substitution (*AES*) by $E_{ij}=w_jAES_{ij}$, where w_j is the jth commodity's expenditure share.) A useful measure of the dispersion of relative prices is

$$D=\tfrac{1}{2}\sum_i w_i^0(r_{it}-L_t)^2 \tag{9}$$

where $r_{it}=p_{it}/p_{i0}$ and w_i^0, respectively, are the price relative and base period budget share for good i, and $L_t=\Sigma_i w_i^0 r_{it}$ is the Laspeyres price index. This measure is a special case of an approximation to the substitution bias derived by John Paulus. The terms $(r_{it}-L_t)$ represent deviations of change in individual relative prices from the average change, the latter measured by the Laspeyres index (a weighted average of price relatives). Thus equation (9) gives a (budget share) weighted sum of squared deviations of relative prices from their weighted average.

The above measures serve as the basis for making inferences about the relative contributions of substitution and of relative price changes in determining whether consumption categories yield a high or low bias in the fixed-weight index. That is, we can compare values of the weighted averages of compensated own-price elasticities and of equation (9)—which I term the "index

[12]The Personal Business and Entertainment groups (along with Recreational Goods) also show the widest range across models in estimated biases, and thus the greatest sensitivity to model specification. For Personal Business, the estimated biases for the *LES*, *GLES*, and indirect addilog, respectively, are 15.8, 15.8, and 8.4 percentage points. For Entertainment, they are 7.4, 6.7, and 11.2 points, and for Recreational Goods, 1.9, 1.6, and 3.1. The most likely reason for the greater range of bias estimates for these groups is that the range of estimated price elasticities is greater than for the other groups (see my paper for a table of elasticities, and also see the discussion below).

TABLE 3—INDEXES OF PRICE DISPERSION AND BUDGET SHARE WEIGHTED AVERAGES OF COMPENSATED OWN-PRICE ELASTICITIES[a]

	Dispersion Index	*LES* Elasticities
Overall	3.5	−.39
I. Food	1.1	−.46
II. Clothing	1.0	−.66
III. Shelter	1.1	−.35
1. Housing and Utilities	0.3	−.18
2. Household Furnishings	0.5	−.43
3. Household Durables	3.8	−.99
IV. Transportation	3.6	−.24
1. Auto Services	2.5	−.19
2. Public Transportation	3.5	−.74
V. Personal Services	10.8	−.00
1. Medical Care	8.2	.23
2. Personal Business	27.7	−.44
3. Personal Care	0.8	.01
VI. Recreation and Entertainment	8.9	−1.0
1. Recreational Goods	4.6	−.63
2. Entertainment	4.3	−1.9

[a]The elasticities were calculated using a base 1958 indifference curve and 1958 prices.

of price dispersion"—with the estimates of the substitution bias from Table 2. The relevant elasticities and indexes of price dispersion appear in Table 3.

It can be seen from Table 3 that the demand for commodities within some groups is more elastic than in others, indicating greater substitution. Furthermore, the high-bias groups (Household Durables, Public Transportation, Personal Business, Recreational Goods, and Entertainment) are all commodity groups which have elasticities in the upper range (with the exception of Personal Business).[13] The only group with comparably high elasticities, but with a low bias, is Clothing (but that group has a low index of price dispersion).

In fact, as indicated in Table 3, the extent of relative price change within the various groups of commodities differs considerably, and the high-bias groups have the largest indexes of price dispersion (Medical Care is again an exception). Note, however, that a high dispersion index does not necessarily imply a high bias (evidence Auto Services).

D. *Alternative Time Periods*

The above analysis was conducted for only one time period, 1958–73. Because the estimated substitution elasticities for all three models were quite stable over time, the choice of time period should only be important to the size of the bias if relative price changes vary from one period to another. The size of the bias should also be related to the length of time that the Laspeyres weights are held constant. Thus it is of interest to examine more than one time period, and to consider the effects of varying the length of time that the weights are held constant.

I examined two five-year intervals within the fifteen-year period used above, 1958–63 and 1968–73—the first characterized by relatively low (5.4 percent) and the second by

[13]The Personal Business group has some atypical characteristics. It has a very high index of price dispersion (see Table 3) which is caused by a radical increase in the price of one of the goods (Brokerage Charges) in the group relative to the prices of the others. This helps explain the high bias estimated by the *LES*, which found demand for that good price elastic, and the lower bias estimated by the indirect addilog (see fn. 12) which found demand for that good inelastic. One could argue that these items represent investment, not consumption, behavior, and do not belong in consumption data (none of these items is in the official *CPI*).

TABLE 4—OVERALL SUBSTITUTION BIAS BY TIME PERIOD AND BY LENGTH OF TIME THAT LASPEYRES WEIGHTS ARE HELD CONSTANT

Period	Change in Overall Prices[a]	Index of Relative Price Dispersion	Substitution Bias with Laspeyres Weights Constant for: 5 years	15 years	25 years
1958–63	5.4	.11	.13	.36	–
1968–73	23.9	.20	.21	.60	.80
1958–73	47.5	3.5	–	1.5	2.5

Note: The base year of the weights used in calculating the Laspeyres indexes is the last year of the period less the length of time the weights are held constant. For example, for the 1958–63 period with weights constant fifteen years, the base year is 1948. The substitution bias is calculated as the difference between the appropriately weighted Laspeyres index and the *LES, CLI* of the same base year.

[a]Shown in percent.

relatively high (23.9 percent) inflation, as measured by the overall Laspeyres price index. A set of four measures was calculated for each five-year period: the amount of *overall* price change (measured by the change in the Laspeyres price index); the amount of *relative* price change (measured by the index of price dispersion described earlier); the substitution bias of a Laspeyres index with weights constant for five years (base year equal to the first year of the period); and the substitution bias of a Laspeyres index with weights held constant for fifteen years (base year ten years earlier than the first year of the period). A fifth measure, the substitution bias with weights constant for twenty-five years was calculated for the 1968–73 period as well as the entire fifteen-year period (1958–73).

The results for the overall indexes are presented in Table 4. Computations for individual commodity groups (available from me on request) follow similar patterns. As indicated in the second column of Table 4, relative prices changed more in the more inflationary (1968–73) period—the overall dispersion index is .20 vs. .11 for the earlier period. As expected, the substitution bias is directly related to the dispersion measure —the total bias (the third and fourth columns) is nearly two-thirds larger in the later five-year period, regardless of how long the weights are held constant. Thus the variation in the measured substitution bias over different intervals depends on the extent of relative price change. In the present case relative prices changed more in the high inflation period than in the low inflation period, although that relationship need not always be true.

A comparison of the substitution bias measures across columns reveals that lengthening the period of time that the Laspeyres weights are held constant considerably increases the size of the bias. In fact, the bias in the earlier, "lower inflation" period with weights constant for fifteen years is greater than that in the later high inflation period with weights constant for five years. Thus, even with a low rate of inflation and small changes in relative prices, the length of time that the Laspeyres weights are held constant has an important effect on the substitution bias.

The practical importance of holding the Laspeyres weights constant for lengthy periods of time can be seen by examining the bias measures in the last row of Table 4. The fifteen-year period used in this analysis was chosen intentionally to approximate the length of time between major weight revisions in the *CPI*. The relatively small bias (0.1 points per year) over this interval might lead one to suggest that the weights in the *CPI* could be held constant for a longer period of time without serious error in measurement of the cost of living. The results in the table indicate, however, that lengthening the time that the Laspeyres weights are held constant from fifteen to

twenty-five years increases the overall bias from 1.5 to 2.5 points, or an increase in the average annual bias from .1 to .17 points. Considerably larger biases emerge in some of the categories (for example, the bias is 7.8 for Household Durables, 9.4 for Public Transportation, and 9.2 for Recreational Goods). The size of these biases, though not overwhelming, does indicate that considerable distortion in the measurement of the cost of living could result from holding the weights in a Laspeyres price index constant for such a long period of time.

The above factors may help explain the difference between the present results and the extremely small biases in measuring the overall cost of living reported in two previous aggregate studies. Goldberger-Gamalestsos and Phlips (using Sanz-Ferrer's estimates) report total biases of .3 index points over an eleven-year period, and .17 points over six years, respectively. These translate into an average annual bias of less than three-hundredths of one point.[14] However, their studies 1) use highly aggregate goods, which are presumably not highly substitutable, and thus ignore substitution among commodities within these aggregate "goods"; 2) use sample periods (1950–61 for Goldberger-Gamalestsos and 1963–69 for Phlips) which were characterized by smaller over-all price changes (and thus possibly smaller relative price changes) than the more recent years covered in the present study; and 3) use a relatively short length of time between the base and comparison periods.

[14]An exception to these small reported biases is a large bias found by Hoa for Australia. However his result is not a valid test of the size of the substitution bias since the proper standard for comparison is a Laspeyres index based on the same data used in the estimation of the true *CLI*, and Hoa compared his estimated *CLI*s to official Australian *CPI*s. In another vein, two recent attempts have been made to provide an estimate of the substitution bias without actually estimating a demand system. Nicholas Noe and George von Furstenberg assume that the consumer's utility function is Cobb-Douglas (one of the most restrictive systems available), and use as parameter estimates values of the expenditure shares in the base period. Their resulting *CLI* (a geometric mean of prices, due to the Cobb-Douglas form) differs only slightly from the *CPI* (an arithmetic mean of prices). The major weakness of the study is the use of such a restrictive utility function as the standard of comparison. They do not make use of less restrictive demand systems, several of which have been estimated by numerous researchers over the past twenty-five years (see Brown and Deaton). P. J. Lloyd, on the other hand, specifies a two-level *CES* utility function with two branches and two goods in each branch. He then *assumes* values for the relative price between the two goods in each branch, and for the budget share and elasticity of substitution parameters—that is, no actual data are employed in his computations. Such an exercise says nothing about what the biases are in practice.

IV. Conclusion

The present study evaluates the amount of substitution bias in a Laspeyres-type price index such as the *CPI*. It is the first attempt to estimate an exhaustive set of cost-of-living subindexes for groups of commodities covering the full range of the consumer budget, as well as an overall cost-of-living index. In addition, the sensitivity of the indexes to specification of functional form is examined by comparing indexes corresponding to three alternative multilevel demand models.

The major results of the study may be summarized as follows:

1) The cost-of-living indexes and subindexes are quite robust with respect to the alternative demand models, although this conclusion is tempered by the restrictiveness of the three models employed.

2) The amount of estimated substitution bias varies considerably among components. The estimated bias of the overall Laspeyres price index is 1.5 index points over the fifteen-year period from 1958 to 1973. The biases of Laspeyres indexes for the groups of commodities range from less than one index point to nearly sixteen points over the period, although more than half of the groups have biases of less than two points.

3) It is clear that the size of the bias depends on the extent of relative price change and the amount of commodity substitution. The groups with the largest biases are found to be characterized by relatively high compensated demand elasticities (indicating considerable substitution) and large changes in relative prices.

4) Increasing the period of time for which the Laspeyres weights are held constant from fifteen to twenty-five years appreciably increases the size of the bias.

Appendix—Groups, Subgroups, and Commodities

I. Food
  1. Food purchased for off-premise consumption
  2. Alcohol purchased for off-premise consumption
  3. Alcohol in purchased meals and beverages
  4. Food in purchased meals and beverages

II. Clothing
  1. Women's and children's clothing
  2. Men's and boy's clothing
  3. Jewelry and watches
  4. Shoes and other footwear

III. Shelter
  1. Housing and utilities
    a. Owner- and tenant-occupied nonfarm dwellings-rental value
    b. Electricity
    c. Gas
    d. Telephone
    e. Other fuel
  2. Household furnishings
    a. Semidurable house furnishings
    b. Cleaning and polishing preparations, and miscellaneous household supplies and paper products
    c. Toilet articles and preparations
    d. Other household furnishings
  3. Household durables
    a. Furniture, including mattresses and bedsprings
    b. Kitchen and other household appliances
    c. China, glassware, tableware, and utensils
    d. Floor coverings
    e. Other durable house furnishings

IV. Transportation
  1. Automobile services
    a. New cars and net purchases of used cars
    b. Tires, tubes, accessories, and parts
    c. Gasoline and oil
    d. Automobile insurance premiums less claims paid
    e. Automobile repairs, greasing, washing, parking, storage and rental
  2. Public transportation
    a. Purchased local transportation
    b. Railway (excluding commutation)
    c. Intercity bus
    d. Airline

V. Personal Services
  1. Medical care
    a. Physicians
    b. Dentists
    c. Ophthalmic products and orthopedic appliances
    d. Drug preparations and sundries
  2. Personal business
    a. Brokerage charges and investment counseling
    b. Bank service charges, trust services, and safe-deposit box rental
    c. Legal services
    d. Funeral and burial expenses
    e. Other personal business
  3. Personal care
    a. Stationery and writing supplies
    b. Postage
    c. Barbershops, beauty parlors, and baths
    d. Tobacco products

VI. Recreation and Entertainment
  1. Recreational goods
    a. Books and maps
    b. Wheels goods, durable toys, sport equipment, boats and pleasure aircraft
    c. Magazines, newspapers, and sheet music
    d. Nondurable toys and sport supplies
    e. Radio and television receivers, records, and musical instruments
  2. Entertainment
    a. Admissions to specified spectator amusements
    b. Commercial participant amusements
    c. Hotels and motels
    d. Other recreation and entertainment

REFERENCES

A. P. Barten, "The Systems of Consumer Demand Functions Approach: A Review," *Econometrica*, Jan. 1977, *45*, 23–52.

Charles D. Blackorby, Daniel Primont, and R. Robert Russell, *Duality, Separability, and Functional Structure: Theory and Economic Applications*, Amsterdam 1977.

______ **and R. R. Russell,** "Indices and Subindices of the Cost of Living and the Standard of Living," *Int. Econ. Rev.*, Feb. 1978, *19*, 229–40.

S. D. Braithwait, "An Empirical Comparison of Alternative Multi-Level Demand Systems for the U.S.," work. paper no. 74, Bureau of Labor Statistics 1977.

A. Brown and A. Deaton, "Surveys in Applied Econometrics: Models of Consumer Behavior," *Econ. J.*, Dec. 1972, *82*, 1145–236.

M. Brown and D. Heien, "The S-Branch Utility Tree: A Generalization of the Linear Expenditure System," *Econometrica*, July 1972, *40*, 737–47.

L. R. Christensen, D. W. Jorgenson, and L. Lau, "Transcendental Logarithmic Utility Functions," *Amer. Econ. Rev.*, June 1975, *65*, 367–83.

L. R. Christensen and M. Manser, "Cost-of-Living Indexes and Price Indexes for U.S. Meat and Produce, 1947–1971," in Nestor Terleckyj, ed., *Household Behavior and Consumption*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 40, New York 1975.

W. E. Diewert, "Homogeneous Weak Separability and Exact Index Numbers," tech. rept. no. 122, IMSSS, Stanford Univ., Jan. 1974.

R. C. Geary, "A Note on a 'Constant Utility Index of the Cost of Living'," *Rev. Econ. Stud.*, No. 1, 1950, *18*, 65–66.

A. S. Goldberger and Theodore Gamaletsos, "A Cross-Country Comparison of Consumer Expenditure Patterns," *Euro. Econ. Rev.*, Spring 1970, *1*, 357–400.

W. M. Gorman, "Separable Utility and Aggregation," *Econometrica*, July 1959, *27*, 469–81.

T. V. Hoa, "Additive Preferences and Cost of Living Indexes: An Empirical Study of the Australian Consumer's Welfare," *Econ. Rec.*, Sept. 1969, *45*, 432–40.

Hendrik S. Houthakker, "Additive Preferences," *Econometrica*, Apr. 1960, *28*, 244–57.

______ **and Lester Taylor,** *Consumer Demand in the United States*, Cambridge, Mass. 1970.

L. R. Klein and H. Rubin, "A Constant Utility Index of the Cost of Living," *Rev. Econ. Stud.*, No. 2, 1948, *15*, 84–87.

L. Lau, "Duality and the Structure of Utility Functions," *J. Econ. Theory*, Dec. 1969, *1*, 374–96.

A. G. Lawrence, "The Bias and the Asymptotic Variance of a Computed True Cost-of-Living Index: The Case of the Klein-Rubin Constant Utility Index," work. paper no. 20, Bureau of Labor Statistics 1974.

P. J. Lloyd, "Substitution Effects and Biases in Nontrue Price Indices," *Amer. Econ. Rev.*, June 1975, *65*, 301–13.

M. E. Manser, "A Note on Cost of Living Indexes for U.S. Food Consumption, 1948–1972," work. paper no. 57, Bureau of Labor Statistics 1975.

N. N. Noe and G. von Furstenberg, "The Upward Bias in the Consumer Price Index Due to Substitution," *J. Polit. Econ.*, Nov./Dec. 1972, *80*, 1280–94.

J. D. Paulus, "Note on: 'The Upward Bias in the Consumer Index Due to Substitution'," spec. stud. paper no. 45, Federal Reserve Board 1974.

Louis Phlips, *Applied Consumption Analysis*, Amsterdam, 1974.

R. A. Pollak, (1971a) "Additive Utility Functions and Linear Engle Curves," *Rev. Econ. Stud.*, Oct. 1971, *38*, 401–14.

______, (1971b), "The Theory of the Cost-of-Living Index," work. paper no. 11, Bureau of Labor Statistics, June 1971.

______ "Generalized Separability," *Econometrica*, May 1972, *40*, 431–53.

______, "Subindexes in the Cost of Living," *Int. Econ. Rev.*, Feb. 1975, *16*, 135–50.

P. A. Samuelson and S. Swamy, "Invariant Index Numbers and Canonical Duality:

Survey and Synthesis." *Amer. Econ. Rev.*, Sept. 1974, *64*, 556–93.

R. Sanz-Ferrer, "Previsions de la consommation privee en Belgique," *Rech. Econ. Louvain*, 1972, *38*, 17–37.

R. Stone, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand" *Econ. J.*, Sept. 1954, *64*, 511–27.

R. H. Strotz, "The Empirical Implications of a Utility Tree" *Econometrica*, Apr. 1957, *25*, 269-80.

______, "The Utility Tree: A Correction and Further Appraisal," *Econometrica*, July 1959, *27*, 482–89.

The Changing Cyclical Behavior of Wages and Prices: 1890–1976

By JEFFREY SACHS*

The persistence of inflation during periods of high unemployment poses the central problem for macro-economic policy in the 1980's. The forecasts of major econometric models for the United States broadly agree that a sustained period of underemployment of resources will be required to markedly reduce the prevailing rate of inflation. Indeed, many economic commentators have surmised that a given level of unemployment now "buys" a smaller reduction in the rate of inflation than in the past. In technical terms, they suggest that the slope of the short-run Phillips curve has declined over time.

In the vast econometric literature on the Phillips relationship, surprisingly little formal analysis has been made of long-term changes in the curve's parameters. Both methodological problems and data limitations make a long-term analysis difficult. Two studies of the secular trend in cyclical wage and price flexibility have recently appeared. The studies use different analytic techniques and reach opposing views on the changing inflation-unemployment tradeoff. In an innovative study of wholesale price behavior in the 1920's and the post-World War II period, Phillip Cagan concludes that "wholesale prices show a smaller decline in the recessions after 1948–49 than formerly," and that "there has clearly been a gradual decline in price response to recessions over the postwar period, except mainly for raw materials prices" (pp. 54–55). Michael Wachter, contrariwise, finds in a study of wages in the post-World War II period that, "a broad range of wage equations reveals the growing cyclical responsiveness of wage inflation" (p.116). The results of course are not directly comparable, for Cagan's focus is on prices while Wachter's is on wages, and the time periods of analysis are different. The discrepancy in conclusions seems to flow, however, from more fundamental differences with the two analyses.

In this paper, two approaches are employed to compare wage and price macrodynamics in the periods 1890–1930 and 1947–76. Both approaches strongly support the hypothesis of a decreasing responsiveness of inflation to changes in aggregate demand. The approach in Section I follows Cagan in analyzing changes in the rates of wage and price inflation from business cycle peaks to troughs. By comparing the decelerations of inflation in pre- and post-World War II recessions of nearly equal magnitude, a rough-and-ready measure of changing price responsiveness is found. In Section II, econometric Phillips curve estimates, akin to Wachter's, are presented. Attention is given to the problems of simultaneous equations bias in standard Phillips curve estimates. A new method of Phillips curve estimation is then described and tested. Not only does the new approach readily yield consistent estimates of the important parameters of the Phillips curve, but it also provides an easy framework for some problems of macro-economic policy.

In Section III, I speculate on the causes for the increase in cyclical wage and price rigidity. It is suggested that two important aspects of the diminishing slope of the short-run Phillips curve are long-term wage agreements and the public's expectations after World War II that monetary and fiscal authorities will intervene to prevent price deflations and unemployment. Theoretical support for these hypotheses is cited, though empirical testing must await further research.

*Harvard University and National Bureau of Economic Research. I gratefully acknowledge the assistance of George Borts, Martin Feldstein, and an anonymous referee.

I

From 1890 to 1976, the American economy experienced nineteen business cycles, as measured by the National Bureau of Economic Research. From 1890 to 1945, fourteen business cycles occurred, and five complete business cycles have transpired since the end of World War II. Following Cagan, simple calculations are made in this section to gauge the responsiveness of wage and price inflation to changes in aggregate economic activity, as indicated by the Bureau cycles. A measure of business cycle severity is used to relate the sizes of wage and price fluctuations to the amplitudes of the cycles.

Modern equations of inflation describe the rate of change of wages and prices according to "equilibrium" and "disequilibrium" components. For some parameter value ϕ, the postulated relationship for wage change at time t is

$$(1) \qquad Dw_t = \phi(L_t^d - L_t^s) + Dw_t^e \qquad \phi > 0$$

where $Dw_t = (1/w)(dw/dt)$, the time rate of proportional change of wages at time t, L_t^d and L_t^s are the demand and supply of labor, and Dw_t^e is the expected rate of wage inflation. Typically, the excess demand for labor, $L_t^d - L_t^s$, is measured by an adjusted rate of unemployment or other index of aggregate activity. The mechanisms of price change may similarly be specified as $Dp_t = \delta(y_t^d - y_t^s) + Dp_t^e$, where y_t is aggregate output.

Whenever Dw^e or Dp^e is fairly sluggish over a business cycle, a comparison of inflation at the cyclical peak and trough allows us to measure approximately the responsiveness of inflation to disequilibrium market conditions. Simply put, for a constant Dw^e, the change in inflation from peak to trough is given by

$$(2) \quad Dw^p - Dw^t = \phi(L^d - L^s)^p - \phi(L^d - L^s)^t$$

where the superscripts indicate peak and trough. If $Dw^p - Dw^t$ falls over time through successive cycles we may conclude that there is an increasing cyclical rigidity of inflation rates (i.e., a decreasing ϕ) when 1) the successive cycles are of similar amplitude, and 2) Dw^e is invariant over the cycle. When Dw^e varies over the cycle, $Dw^p - Dw^t$ will measure changes both in equilibrium and disequilibrium components.

Let us now turn to the historical evidence, using price and wage indexes for 1890–1929 and 1947–76. Throughout this paper, the period of the Great Depression and World War II are excluded from the analysis. The period of the 1930's is skipped because of well-known perversities in the wage and price dynamics of the period.[1] The years of World War II similarly must be eliminated because of the extensive administrative control of wages and prices during the period. The empirical results displayed in this section are for the *WPI* and indexes of average hourly compensation.[2] The calculations have also been made for a number of other price and wage series, with very similar results. For the sake of brevity, these additional results are not reported, but are available from me upon request.

[1] With unemployment hovering above 20 percent in 1933, prices and wages stopped a four-year fall, and actually began to rise at significant rates. The Wholesale Price Index (*WPI*) rose an average of 6.7 percent per year from 1933 to 1937 while the National Income Accounts wage measure for manufacturing workers increased an average of 5.9 percent per year. Wachter notes that: "...there is evidence that government measures to spur recovery were behind the upward movement in wages. The spurt in the early 1930s is generally attributed to the National Industrial Recovery Act, and Wagner Act and the subsequent growth of unions may have been responsible for that of the late 1930s," (p. 155).

Milton Friedman and Anna Schwartz reach similar conclusions. In accounting for the large rise in the *WPI*, they cite "the explicit measures to raise wages and prices undertaken with government encouragement and assistance, notably, NIRA, the Guffey Coal Act, the agriculture price-support program, and the National Labor Relations Act" (p. 233). Because of these complex institutional changes, the Great Depression is not in the interval of analysis.

[2] Close comparability in wage data for the two periods requires focusing on the manufacturing sector, given data limitations in the earlier period. For the recent period, I have compared Phillips curves estimated with manufacturing sector and economy-wide wages as the dependent variable, finding very little difference in cyclical responsiveness.

TABLE 1—PRICE AND WAGE CHANGES OVER THE BUSINESS CYCLE

		Change, Wholesale Price Index[a]			Change, Compensation Per Hour, Manufacturing[a]		
Year Before Peak to Peak	Peak to Trough	Year Before Peak to Peak (1)	Peak to Trough (2)	(2)−(1)	Year Before Peak to Peak (3)	Peak to Trough (4)	(4)−(3)
1892–93	1893–94	2.2	−10.7	−12.9	4.9	−7.2	−12.1
1894–95	1895–97	2.0	−2.4	−4.4	−3.1	1.3	4.4
1898–99	1899–1900	7.3	7.1	−.2	5.6	1.4	−4.2
1901–02	1902–04	6.4	.6	−5.8	7.3	3.1	−4.2
1906–07	1907–08	4.8	−3.6	−8.4	3.5	−4.9	−8.4
1909–10	1910–11	4.2	−8.3	−12.5	6.2	1.3	−4.9
1912–13	1913–14	1.1	−2.2	−3.3	7.1	−.2	−7.3
1917–18	1918–19	10.9	−5.5	−16.4	27.8	13.0	−14.8
1919–20	1920–21	10.8	−46.0	−56.8	15.0	−12.4	−37.9
1922–23	1923–24	3.9	−2.7	−6.6	10.0	3.3	−6.7
1925–26	1926–27	−3.2	−4.5	−1.3	.9	.9	0.0
1947–48	1948–49	7.9	−5.1	−13.0	8.6[b]	3.1[b]	−5.5[b]
1952–53	1953–54	−1.3	.2	1.5	5.5	4.3	−1.2
1956–59	1957–58	2.8	1.4	−1.4	5.7	4.3	−1.4
1959–60	1960–61	.1	−.4	−.5	3.9	2.9	−1.0
1968–69	1969–70	3.8	3.5	.3	6.3	6.5	.2
1972–73	1973–75	12.3	13.1	.8	7.0	9.9	2.9

Sources: Compensation per hour, 1890–1927, is from the Rees series, expressed in nominal terms. Compensation per hour, 1952–75, is the *BLS* series on Average Hourly Compensation in Manufacturing. For both series, see the Appendix.

Note: Percent Change $\text{Index}_t = \{Log(\text{Index}_t) - Log(\text{Index}_{t-i})\}/i$, where i is the number of years in the interval.

[a] Percent change, annual rate.

[b] For 1947–49, economy-wide compensation per hour is used. The *BLS* series on Average Hourly Compensation in Manufacturing begins in 1950.

Calculations of annual inflation rates are made for seventeen cycles. In Table 1, I report the rates of inflation before and after business cycle peaks for the price and wage series. Under each index, the third column is the difference of the inflation rates from peak to trough of the cycle. The evidence strongly supports the role of aggregate activity on price and wage change. For the *WPI*, the rate of inflation declined during every downturn from 1890 to 1948, and for three of the five recessions after 1948. Wages, similarly, decelerated during every contraction from 1890 to 1961, excluding the 1895–97 period.

In complete accord with Cagan's results for the 1920's, the *WPI* calculations support the hypothesis that prices have become more cyclically rigid over time (though of course I have as yet made no attempt to control for cyclical severity). Almost every contraction from 1890 to 1927 produced a sharper deceleration in price change than did later recessions. Only the deceleration in 1949 is of similar magnitude with the earlier cycles. It is plausible that changing weights in the construction of the *WPI* might account for some of the apparent increase in rigidity. Note, for example, that cyclically flexible agriculture and food prices composed 43.06 percent of the *WPI* in 1926, and only 20.95 percent in 1970. Cagan, for a more limited time period, reestimated the shifts in cyclical inflation using a fixed weighting scheme, and found very little change in the results (pp. 99–100). As a further check, calculations (not shown) were made for the subindex of industrial goods, available from Bureau of Labor Statistics (*BLS*) since 1913. The same perceptible decline in price flexibility is evident.

In Table 1, there is also a discernible trend toward greater cyclical inflexibility of wage change. While nominal wage growth slowed in ten of the eleven cycles before 1930, wage change has in fact accelerated in

TABLE 2—WAGE AND PRICE BEHAVIOR DURING BUSINESS CYCLES: CLASSIFIED BY CYCLICAL SEVERITY, 1890–1976

Peak-Trough	Change in Price Inflation	Change in Wage Inflation	$\Delta\ GAP \times 100$
Mild Contractions			
1893–94	−12.9	−12.1	−7.5
1895–97	−4.4	4.4	−8.5
1899–1900	−.2	−4.2	−1.7
1910–11	−12.5	−4.9	−8.2
1923–24	−6.6	−6.7	−9.2
1926–27	−1.3	0.0	−3.7
1953–54	1.5	−2.1	−9.7
1960–61	−.5	−1.0	−3.9
1969–70	.3	.2	−8.4
Moderate Contractions			
1902–04	−5.8	−4.2	−11.0
1913–14	−3.3	−7.3	−10.5
1948–49	−13.0	−5.5	−10.2
1957–58	−1.4	−1.4	−11.4
Strong Contractions			
1907–08	−8.4	−8.4	−21.0
1918–19	−16.4	−14.8	−17.4
1920–21	−56.8	−37.4	−25.9
1973–75	.8	2.9	−15.8

Source: Wage and price data from Table 1.

Note: The output gap variable is calculated by regressing *log* (Industrial Production Index) on time and the rate of unemployment. The rate of unemployment is then set at its average value, and fitted values of *log* (Industrial Production Index) are interpreted as trend values. These trend measures are subtracted from actual values, to obtain GAP_t. The calculations were performed separately for the pre- and post-World War II period. The cycles have been grouped by severity, according to ΔGAP, where $\Delta GAP = GAP^p - GAP^t$, the change in GAP from peak to succeeding trough. The grouping is Mild contraction: $-10 \leq \Delta GAP \times 100 \leq 0$; Moderate contraction: $-15 < \Delta GAP \times 100 < -10$; Strong contraction: $\Delta GAP \times 100 < -15$.

the last two downturns. Furthermore, the magnitudes of deceleration before 1930 are in general far larger than the measures for 1947–60, where wage change did in fact slow down.

It is important to compare cycles of similar magnitude for testing changes in cyclical wage and price flexibility. Part of the apparent increase in rigidity is due to the relative mildness of most of the post-1945 recessions. The Bureau rankings of business cycle severity exist for 1920–76. I use these rankings as a "benchmark" for a measure covering the entire 1890–1976 period.[3] In particular, a measure of severity is calculated from the percentage deviation of industrial output from its trend value at business cycle peaks and troughs. The change in this output "gap" from peak to trough gauges the amplitude of the cycle. For the years 1920–76, the gap measure and the Bureau severity index yield almost identical rank orderings.

In Table 2, the cycles are arranged according to the gap criterion. The precise classification method is described in the table. The evidence is rather striking. For mild contractions, downward price flexibility seems to have ended with the pre-World War II period. For moderate and severe contractions, similarly, the response of wages and prices has fallen significantly since 1950.

II

The methods of Section I, though convincing in their illustration of a trend toward decreasing cyclical changes in inflation, fail to yield actual estimates of the

[3] See National Bureau of Economic Research, especially *53rd Annual Report* (1973, p. 18).

Phillips curve parameters. It is shown below, however, that traditional econometric attempts to estimate changes in the inflation-unemployment tradeoff may be flawed. In this section, I describe a simple method for obtaining consistent estimates of the Phillips curve coefficients. These estimates add further strong support to the hypothesis of a decline in the short-run Phillips curve slope.

In the standard Phillips curve analysis, (1) is estimated in the form

$$(3) \quad Dw_t = \beta_0 + \beta_1 RU_t^{-1} + \beta_2 y(L)Dp_{t-1} + e_t$$

Here, $y(L)$ denotes a polynomial of the lag operator, and RU_t^{-1}, the inverse of the unemployment rate, substitutes for $L_t^d - L_t^s$ in (1). The term RU_t is conventionally entered in inverse form as a reflection of the assumption of a non-linear, decreasing response of inflation to increasing levels of unemployment. Lagged values of the unemployment rate may also be entered. Expected inflation is measured by $y(L)Dp_{t-1}$. The somewhat arbitrary nature of the specification $Dp_t^e = y(L)Dp_{t-1}$ is now well-known, after the justified criticism of rational-expectations theorists.[4] However, for our purposes, the distributed-lag approach may be warranted. In an economy (for example, pre-1930) where the macro-economic structure is not well understood, much less precisely estimated in econometric work, it may be "economically rational" (see Edgar Fiege and Douglas Pearce) to use forecasts of future inflation based upon lagged values of price change. Indeed, when parameter values of structural equations are not known, adaptive or error-learning procedures of forecasting inflation are often optimal forecasting procedures.

Direct estimates of (3) were made for the pre- and post-World War II period, with varying assumptions about the unemployment variable and the structure of $y(L)Dp_t$. All of the regressions reported in this paper use the measure of output gap in industrial production as the disequilibrium variable. For a number of reasons, it is believed that output gap should be preferred to the Lebergott series for pre-1930 labor force unemployment.[5] Various attempts were made to enter lagged values of GAP_t in regression estimates of (3). When GAP_{t-1} was included along with GAP_t, the coefficient of the lagged variable was small and never statistically significant, while the current variable remained significant. When GAP_{t-1} was entered instead of GAP_t, the coefficient was almost always insignificant.

The estimates for (3) are presented in Table 3. All estimates show a large decline in β_1 from the first to the second period, that is, the short-run Phillips curve is flatter in the second period, in conformity with the findings of Section I. In all of the equations, the short-run Phillips curve slope estimate is of the right sign and is statistically significant. For the pre-World War II period, $\hat{\beta}_1$ varies between .4 and .53, while for the post-World War II era, $\hat{\beta}_1$ lies between .07 and .12. The parameter of adaptive expectations in Dp_t^e, λ, is not well measured, and has the wrong sign in regression 3.

In his analysis, Wachter reached very different conclusions regarding the change of β_1 over time. Wachter concluded that the short-run Phillips curve has become increasingly steep. His analysis concerned the 1954–75 period, while the present paper compares the pre-World War II and post-World War II periods, so that the results are not strictly contradictory. Yet it is still worthwhile to ask whether Wachter's methodology might explain the variance of the results. He estimates a modified form of (2):

$$(4) \quad Dw_t = \beta_0 + \beta_1 RU_t^{-1} + \beta_2 y(L)Dp_{t-1} + \beta_3 RU_t^{-1} \cdot TIME$$

[4] See, for example, Robert Lucas.

[5] The Lebergott series is subject to large errors. The only available benchmarks for pre-1930 unemployment are the decennial census data of 1900, 1910, and 1920. All other years must be interpolated. Random error in the unemployment rate series would tend to bias toward zero the estimated coefficient on RU_t. Indeed, using an instrumental variable for RU_t substantially increased the magnitude of the unemployment rate coefficient, as would be predicted with errors-in-variables. Importantly, in the regressions I analyzed, the results for shifts in the Phillips curve hold for *both* RU_t and GAP_t as explanatory variables.

TABLE 3—PHILLIPS CURVE ESTIMATION, 1890–1976
(Annual Data)

Period	β_0	β_1	β_2	Dp^e	Post-World War I Dummy Variable	R^2	DW	$\hat{\rho}$
1. 1897–1929	.002	.53	constrained	$\lambda=.41$	.057	.71	1.13	
	(.24)	(6.4)	to 1	(4.3)	(2.2)			
2. 1954–76	.025	.087	constrained	$\lambda=.54$		.99	1.62	
	(11.4)	(2.72)	to 1	(5.66)				
3. 1897–1929	.013	.40	.49	$\lambda=-.40$	.09	.79	1.19	
	(1.9)	(5.56)	(5.56)	(3.48)	(4.1)			
4. 1952–76	.020	.071	1.19	$\lambda=.59$		.99	1.85	
	(3.38)	(1.65)	(5.61)	(5.86)				
5. 1894–1929	.019	.43	$\sum_1^3 \lambda_i=.46$		.07	.79	1.76	.14
	(1.48)	(5.34)			(3.23)			(3.45)
6. 1952–76	.02	.114	$\sum_1^3 \lambda=1.14$			.997	1.88	−.35
	(5.78)	(3.53)						(1.88)

Notes: The dependent variable is percentage change in Average Hourly Compensation in Manufacturing, as described in Table 1. The variable GAP_t is used in place of RU_t^{-1} shown in (2), for reasons described in Section II. In regressions 1–4,

$$Dp_t^e=(1-\lambda)\sum_{i=0}^{3}\lambda^i Dp_{t-q-i}+\lambda^4 Dp_{-5}$$

where Dp_t is the percentage rate of change of the *GNP* deflator, described in the Appendix. In regressions 5–6, three lagged values of Dp were entered in unconstrained form as the measure of Dp_t^e. The Post-World War I Dummy Variable is included to account for the massive deflation of 1921. Post-World War I Dummy = 1 in 1919 and 1920, and −1 in 1921. Regressions 1–4 were estimated using non-linear squares. Regressions 5–6 were estimated with the Cochrane-Orcutt correction for serial correlation, given by $\hat{\rho}$. The numbers in parentheses are t-statistics.

Since $dDw_t/dRU_t^{-1}=(\beta_1+\beta_3\cdot TIME)$, Wachter avers that a positive β_3 should indicate a secularly increasing responsiveness of inflation to unemployment, while a negative β_3 should prompt the reverse conclusion. For 1954:I to 1975:II, he finds $\hat{\beta}_3>0$. Wachter errs, however, in letting only the coefficient on RU^{-1} change over time. If, in the true model, either β_0 or β_2 but *not* β_1 increases over time, then estimating (4) will tend to give an upward biased estimate of β_3.[6] In fact, it is far more likely that β_2, the coefficient on the distributed lag on prices, has been increasing while β_1 has been falling. I show below the results of estimating Wachter's equation, without β_3, for different terminal dates. It seems clear that β_1 has not significantly increased, when analyzed by a method that allows all parameters to change:[7]

Time Period	$\hat{\beta}_1$	$\hat{\beta}_2$
1954:I–1965:IV	.024	.31
	(1.3)	(.58)
1954:I–1968:IV	.017	.31
	(1.9)	(.71)
1954:I–1976:IV	.011	.87
	(2.1)	(5.11)

[6] Suppose, for example, that the curve's intercept but not slope has been increasing over time, so that $Dw_t=\beta_0+\lambda_1 TIME+\beta_1(RU^{-1})_t+\beta_2 y(L)Dp_{t-1}, \lambda_1>0$, is the true model. If equation (4) is estimated instead, it is easily shown that $plim(\beta_3)=\lambda_1 reg(TIME, GAP\cdot TIME|y(L)Dp_{t-1}, RU_t^{-1})$, where the term $reg(\cdot)$ is the multiple regression coefficient of $GAP\cdot TIME$ in a regression of $TIME$ on $GAP\cdot TIME$, $y(L)Dp_{t-1}$, and RU_t^{-1}. Thus, the estimate of β_3 would tend to be positive even though β_3 is in fact zero.

[7] The equation $Dw_t=\beta_0+\beta_1 RU_t^{-1}+\beta_2 y(L)Dp_{t-1}$ is estimated for quarterly data. The term w_t is the average hourly earnings in manufacturing; RU_t is the unemployment rate for males age 25+; p_t is the nonfarm deflator, and $y(L)$ is a third-degree polynomial distributed lag of length sixteen quarters, unconstrained at both ends.

The estimates that we have so far examined in Table 3 and Wachter's estimates may be marred by a bias introduced in regressing wage changes on price changes. Most current econometric price equations confirm that prices are well described as markups over standard unit labor costs. Price changes approximately equal wage changes less trend productivity growth, Dq^t. Rewriting (3) with $Dw_{t-1} - Dq^t = Dp_{t-1}$, we have

$$(5) \quad Dw_t = (\beta_0 - \beta_2 Dq) + \beta_1 RU_t^{-1} + \beta_2 y(L) Dw_{t-1} + e_t$$

Now, if the e_t is a serially correlated process, $y(L)Dw_{t-1}$ will be correlated with e_t, and estimates of (5) will be biased. One standard method for dealing with this problem is to assume a particular form for the process e_t, and to make maximum likelihood estimates for (5) using the process explicitly. This approach is taken in equations (5)–(6) of Table 3, where it is assumed that $e_t = \rho e_{t-1} + u_t$, and u_t is independently, identically distributed. Note that $\hat{\rho}$ is statistically significant in both of these equations. With the autocorrelation correction, β_1 remains larger in the earlier period.

There is no particular reason beyond convenience to postulate this specific autoregressive process. With three standard assumptions, however, it is possible to skirt the statistical difficulties of the usual analysis. The assumptions of 1) adaptive inflationary expectations, 2) no long-run inflation-unemployment tradeoff, and 3) prices as markups over standard unit labor costs, lead to the following simple model (writing GAP_t for RU_t^{-1}):

$$(6) \quad Dw_t = \beta_0 + \phi(GAP)_t + Dp_t^e + e_t$$

$$(7) \quad Dp_t^e = (1-\lambda) Dp_t + \lambda Dp_{t-1}^e$$

$$(8) \quad P_t = (1+m) SULC$$

(standard unit labor cost)

If trend labor productivity growth is at a constant rate Dq, then (8) may be rewritten as $Dp_t = Dw_t - Dq^t$. Notice that (7) may be rewritten as $Dp_t^e = (1-\lambda) Dp_t/(1-\lambda L)$. Using this expression with (6) and (8) we may derive

$$(9) \quad Dw_t - Dw_{t-1} = \Delta Dw_t = \frac{(1-\lambda)(\beta_0 - Dq)}{\lambda} + \frac{\phi}{\lambda} GAP_t - \phi GAP_{t-1} + \frac{e_t}{\lambda} - e_{t-1}$$

Through a transformation of (6) we have thus been able to eliminate the lagged wage terms from the estimated equation. Under the assumption that GAP_t is exogenous, the estimation of (9) by maximum liklihood or ordinary least squares will give consistent estimates of the parameters. Finally, note that we may find the "natural rate" of the output gap by setting $\Delta Dw_t = 0$, and $GAP_t = GAP_{t-1}$. We find $GAP^{NR} = (\beta_0 - Dq)/-\phi$. The equation to be estimated becomes

$$(10) \quad \Delta Dw_t = \frac{-\phi}{\lambda}(1-\lambda) GAP^{NR} + \frac{\phi}{\lambda} GAP_t - \phi GAP_{t-1} + \frac{e_t}{\lambda} - e_{t-1}$$

How plausible are the assumptions underlying the present model? Equation (10) suggests an historical consistency check of the model. We can see from (10) that $(GAP_t - GAP_{t-1}) > (1-\lambda)(GAP^{NR} - GAP_{t-1})$ implies ΔDw and $\Delta Dp > 0$. Thus, whenever aggregate activity is rising (i.e., $GAP_t - GAP_{t-1} > 0$) *and* employment or output in period $t-1$ is above the equilibrium level, inflation should intensify ($\Delta Dw > 0$). By similar argument, whenever output is below its long-run equilibrium level, and output is falling, inflation should be decelerating. In eleven of the years since 1893, the economy experienced increasing output relative to potential during a period of already high employment ($GAP_{t-1} > GAP^{NR}$). In ten of these years, wage inflation accelerated as predicted. In nine years of the periods 1893–1929 and 1948–75, the economy was characterized by low and falling levels of aggregate activity. In six years, wage inflation showed a declining rate.[8] The accelerationist property is justified.

[8]In years 1894, 1896, 1900, 1921, 1960, 1961, 1971, 1974, and 1975, $GAP_t < GAP_{t-1} < 0$. In six of nine years, wage inflation decelerated as expected: 1894, 1900, 1921, 1960, 1961, and 1971. In years 1902, 1906,

TABLE 4—PHILLIPS CURVE ESTIMATION, 1893–1975, EQUATION (11)

Period	Dependent Variable	$\hat{\phi}$	$\hat{\lambda}$	$\hat{GAP}^{NR}$	Post-World War I Dummy Variable	SE	DW
1894–1926	$\Delta Dw_t{}^a$	.50	.63	.02	.10	.049	1.92
		(5.36)	(5.82)	(.93)	(3.42)		
1895–1929	$\Delta Dw_t{}^b$	.36	.53	.014	.015	.053	2.19
		(3.57)	(3.88)	(.51)	(.48)		
1895–1929	$\Delta Dw_t{}^d$	.31	.68	.033	.087	.048	2.13
		(3.58)	(3.42)	(.57)	(3.03)		
1950–75	$\Delta Dw_t{}^a$	.14	1.31	.06		.012	2.26
		(2.80)	(2.83)	(.53)			
1952–75	$\Delta Dw_t{}^b$	.038	4.37	.011		.015	1.94
		(.53)	(.16)	(.098)			
1950–75	$\Delta Dw_t{}^c$	.13	.997	−3.03		.014	2.47
		(2.14)	(2.48)	(.06)			
1950–75	$\Delta Dp_t{}^d$	.026	.41	−.10		.018	2.08
		(.32)	(.38)	(.51)			

Notes: For all data, see descriptions in Appendix; $\Delta DX_t = (\log X_t - \log X_{t-1}) - (\log X_{t-1} - \log X_{t-2})$, SE is standard error of the regression. All regressions use non-linear least squares estimation.

[a] w_t is Average Hourly Earnings in Manufacturing.

[b] w_t is Average Hourly Compensation in Manufacturing.

[c] w_t is Average Hourly Compensation, economy wide.

[d] p_t is GNP Deflator.

Equation (10) may be estimated by ordinary least squares (OLS) or by a maximum likelihood procedure (non-linear least squares (NLS)). With OLS we estimate:

$$(11) \quad \Delta Dw_t = \zeta_0 + \zeta_1 GAP_t + \zeta_2 GAP_{t-1} + \mu_t$$

Using the estimates ζ_1 and ζ_2 we may obtain consistent estimates of the underlying parameters ϕ and λ. Note that $plim(\hat{\zeta}_2) = -\phi$, and $plim(-(\hat{\zeta}_2/\hat{\zeta}_1) = -plim(\hat{\zeta}_2)/plim(\hat{\zeta}_1) = \lambda$. In addition, since $\zeta_0 = (-\phi/\lambda)(1-\lambda)GAP^{NR}$, we can also obtain a consistent estimate of GAP^{NR}. By use of the Fieller Bound technique, confidence intervals may be calculated for the point estimates of the underlying parameters (see Zvi Griliches, pp. 32-33). With NLS we may directly estimate the underlying parameters of the model.[9] Because (11) is exactly identified, that is, there is a one-to-one mapping from $(\zeta_0, \zeta_1, \zeta_2)$ to $(GAP^{NR}, \phi, \lambda)$, the OLS and NLS estimates of the underlying parameters are identical.

Table 4, presents a summary of estimates of the model. The results are encouraging. In all of the regressions, the coefficients are of expected sign, and are usually statistically significant. In general, the estimates of ϕ are close to the estimates of β_1 shown in Table 3; again, ϕ declines substantially from the first period to the second. The estimation of Table 4 was made for additional time periods and different wage and price indexes, with little change in the results. For all of the regressions, we cannot reject the hypothesis that $GAP^{NR} = 0$. This is an appealing result. Because GAP_t is constructed as a measure of the deviation of output from trend, the result suggests that the long-run equilibrium value of output (GAP^{NR}) is equal to its trend value. On average, over extended periods, the economy is in equilibrium. Note, finally, that there is a little evidence that λ has increased over time. If so, the mean lag of past inflation in forecasts of future inflation $(\lambda/1-\lambda)$ has also lengthened over time. While λ is a crucial parameter for macro-economic policymaking, as shown below, it is not precisely estimated by the equations of Table 4.

1910, 1913, 1916, 1926, 1929, 1951, 1953, 1966, and 1968, $GAP_t > GAP_{t-1} > 0$. In all but one year (1953), wage change increased as predicted.

[9]Note that u_t in (11) is equal to $e_t/\lambda - e_{t-1}$, from (11). There is no more nor less reason to believe that u_t is identically independently distributed (*iid*) than to believe that e_t is *iid*. The coefficient estimates from NLS (or OLS) regression of (12) will be consistent whether or not u_t is *iid* though efficiency, and consistency of the estimates of the standard errors require that u_t be *iid*.

If ϕ and λ can be considered as exogenous to the policymaker (the discussion of Section III indicates that ϕ and λ change with varying macro-economic policies), we can describe some effects of the shifts of the Phillips curve for policy by a simple formula. Consider a convenient (and highly stylized) one-parameter description of policy. For any output GAP_{t_0} in year t_0, policymakers choose $\delta \cdot GAP_{t_0}$ as their output target in year t_1. Appropriate monetary and fiscal policies are followed so that a constant proportion $(1-\delta)$ of the deviation of output from its equilibrium is removed each year. Then it can be shown that:[10]

(12)

$$Dp(\text{steady state}) = Dp_{t_0} + \frac{\phi GAP_{t_0}(\delta-\lambda)}{\lambda(1-\delta)}$$

Thus, starting in a recession ($GAP_{t_0}<0$), the following conditions imply a high steady-state rate of inflation: 1) a rapid recovery (low δ); 2) long lags in expectations (high $-\lambda$); and 3) a low short-run Phillips curve slope (small ϕ). Given the estimates of falling ϕ and rising λ, we can understand the policy difficulty of recovering from the 1974–75 recession.

Two further points are worthy of mention. First, the path of constant inflation is followed by setting $\delta=\lambda$, as shown in equation (12). The magnitude of the short-run Phillips curve slope is completely irrelevant in determining this path, under assumptions of adaptive expectation! Second, while a low Phillips curve slope and long lags in expectations are very undesirable during recessions, these conditions are *most* desirable during booms. A low slope and long mean lag permit an economy to operate with output in excess of the natural rate for an extended period without a serious acceleration of inflation. Wachter has offered evidence of precisely this phenomenon. Since 1950, Wachter argues, there is a continuing pattern of "smaller first-year upward responses of wages to tight market conditions" (p. 153).

In interpreting (12) it has been assumed that ϕ is independent of the policy parameter δ. In fact, changes in δ over long periods of time may be a major source of long-term movements of ϕ. As will be shown in the next section, the Phillips curve may not be invariant to policy choices; and with ϕ a function of δ, (12) would not be the true "policy menu."

III

We have thus far seen how the parameters of the Phillips curve have changed over time, and how these changes have implications for economic policy. Our understanding of inflation-unemployment interactions would be considerably enhanced by a detailed historical discussion of how and why these changes emerged. Unfortunately, given the immense structural shifts that have characterized the *U.S.* economy since 1900, the historical exegesis is a large task, and one that is beyond the scope of this paper. Complex changes in product and labor markets, such as increasing concentration, higher ratios of value-added per shipment, increased unionization, and the large increases in investment in human capital, have all played a role in the decreased cycli-

[10] For a policy rate δ, we can easily calculate the steady-state conditions for an initial GAP and rate of inflation. Since $Dp_t - Dp_{t-1} = (\phi/\lambda)GAP_t - \phi GAP_{t-1}$ (from (11) setting $e_t = e_{t-1} = GAP^{NR} = 0$):

(a)

$$\sum_{t=1}^{T}(Dp_t - Dp_{t-1}) = \sum_{t=1}^{T}\left(\frac{\phi}{\lambda}GAP_t\right) - \sum_{t=1}^{T}\phi GAP_{t-1}$$

But $GAP_t = \delta^t GAP_{t0}$, by assumption. Replacing this relation in equation (a) we find

(b)

$$\sum_{t-1}^{T}(Dp_t - Dp_{t-1}) = \frac{\phi}{\lambda}GAP_{t0}\sum_{t=1}^{T}\delta^t - \phi GAP_{t0}\sum_{t=0}^{T-1}\delta$$

or, simplifying,

(c)

$$Dp_T - Dp_0 = \frac{\phi}{\lambda}GAP_{t0}\cdot\frac{(\delta-\delta^{T+1})}{(1-\delta)} - \phi GAP_{t0}\frac{(1-\delta T)}{(1-\delta)}$$

To find the steady-state condition, take the limit with respect to T of (c), and rearrange:

$$Dp_{ss} = Dp_{t0} + \frac{\phi GAP_{t0}(\delta-\lambda)}{\lambda(1-\delta)}$$

cal response of wage and price inflation. In this section I shall discuss two less analyzed, though probably very important, contributors to the Phillips curve shift.

First, the emergence of countercyclical macro-economic policy since World War II has probably changed the cyclical behavior of wage and price setters. Martin Baily has recently argued that growing expectations of countercyclical macro-economic policy have smoothed the cyclical adjustments of production and employment in the private sector. It is also likely that such expectations have smoothed the cyclical movements of wages and prices. I have demonstrated this theoretically in my earlier paper, and will outline the main argument below.

A second source of cyclical rigidity is probably the spread of long-term explicit and implicit contracts. It is well-known that union wages are less responsive cyclically than nonunion wages (compare H. Gregg Lewis, Robert Flanagan). The usual explanation points to the use of long-term wage agreements in the union sector. Over my period of study, unionization has spread dramatically, as has the average duration of collective bargaining agreements. This trend could well result in secular increases in aggregate wage rigidity. Note that the spread of long-term contracts might itself result from the stability engendered by active macro-economic policy.

One explanation of the Phillips curve is that workers cut wages when there is high unemployment because of the expectation of continued low demand for labor, at given nominal wages, in the future. To the extent that countercyclical policy breaks the link between current unemployment and the expectation of continuing low aggregate (nominal) demand, today's unemployment will not induce wage cuts in contracts for succeeding periods. In a model of macroeconomic response to a supply shock, Edmund Phelps has succinctly stated this view:

> Suppose that wage setters expect the central bank to accommodate the supply shock by adjusting the money supply and thus the price level in such a way as to hold invariant the quantity of labor that will be demanded by firms at the pre-existing money wage W_0.... If they know they hold these beliefs in common, then their "rational expectation" is that the pre-shock money wage will equilibrate the labor market as it did before the shock. Each firm will expect the other firms to maintain their wages and it will do the same. [p. 209]

In Phelps' case, the slope of the short-run Phillips curve is zero, for unemployment does not induce any wage deflation.

In a more general model of wage setting in the presence of activist policy, I have shown that the slope of the statistical Phillips curve depends on 1) the extent of countercyclical policy, and 2) the degree to which unanticipated changes in the money stock are countered by the monetary authority in succeeding periods.[11] Simple regressions describing money supply growth indicate that both supply characteristics have changed in the direction tending to reduce the Phillips curve slope.[12] In the period 1895–1929 there is no evidence of a countercyclical response of the money supply. For 1952–75, the regression indicates that the money stock is raised 1 percent above trend in the year following a 10

[11] Condition 1) is described by Phelps. Condition 2) results from the fact that any period's GAP_t is in part due to unexpected movements of the money stock, M_t. If M_t falls below anticipated levels, output will drop. The decline in M_t reduces expectations of future full-employment nominal wages only if the shortfall is permanent (i.e., as long as the central bank does not act to "correct" the shortfall in M_t by reflating next period). To the extent that the drop in money growth is not counteracted, $GAP_t<0$ will be followed by a reduction in $w_{t+1}-w_t$.

[12] The regressions relate money growth to lagged money growth and lagged output gap, in an equation similar to Robert Barro's. For $\Delta M_t=\alpha_0+\alpha_1\Delta M_{t-1}+\alpha_2\Delta M_{t-2}+\alpha_3\Delta M_{t-3}+\alpha_4 GAP_{t-1}+u_t$, $u_t=\rho u_{t-1}+e_t$ we find

	α_0	α_1	α_2	α_3	α_4	ρ	R^2
1895–1929	.07	−.31	−.11	.27	.06	.17	.605
	(2.54)	(1.4)	(.62)	(1.6)	(.42)	(1.02)	
1952–75	.09	.55	−.20	.56	−.11	−.53	.49
	(1.05)	(2.45)	(.82)	(2.78)	(1.48)	(2.99)	

where ΔM_t is $log(M_2)_t-log(M_2)_{t-1}$, for the M_2 definition of the money stock. Sources of M_2 are given in the Appendix.

percent industrial output shortfall, and more in following years. Moreover, there is evidence of significant *negative* serial correlation in the residuals of the money supply equation for 1952–75, suggesting that monetary authorities now act to correct partially for unexpected money movements.

The presence of long-term contracts fixing nominal wage growth also reduces the short-run response of aggregate wages to cyclical fluctuations, for two reasons. Most directly, wages fixed by earlier contract may be unable to react at all in the short-term to current, unexpected cyclical developments. This phenomenon is clearly evident in Flanagan's recent study of union-nonunion wage differentials. Comparing the contract (union) sector with the noncontract sector, Flanagan writes:

> That differences in the cyclical sensitivity of average union and non-union wage changes exist is clear in the post-war data. However, it is also clear that first year negotiated wage changes are almost as sensitive to labor market pressures as non-union wages. *Most of the inertia in negotiated wages is a by-product of multi-year labor agreements.* [p. 673, emphasis added]

Second, I have shown in my earlier paper that wages determined in the *noncontract* market (assumed to clear continuously) will show smaller cyclical fluctuations the larger is the sector of labor covered by long-term agreements. Basically, the larger is the contract sector the smaller is the aggregate price disturbance transmitted to the noncontract sector following an aggregate demand shock. The disturbance is absorbed in output fluctuations in the contract sector rather than in aggregate price fluctuations. Assuming that the noncontract labor market is cleared at a given real wage, the reduced aggregate price fluctuations result in smaller wage fluctuations in that sector. These results depend on low intersectoral mobility of labor over the cycle.

Thus the tremendous increase in duration and coverage of collective bargaining agreements are probably important forces behind the Phillips curve shift between the two periods. As late as 1948, the great majority of all wage agreements were of one year duration; by 1972, most contracts were written for three years.[13] And the contracts now cover a larger portion of the work force. In the manufacturing sector, for instance, only 11.6 percent of production workers were organized in 1910, while by 1973, approximately 49 percent of manufacturing production employees were organized by labor unions. Economy wide, 5.8 percent of the civilian labor force belonged to unions in 1910, while 23.4 percent belonged in 1970.[14]

The two explanations for increasing cyclical wage rigidity, activist macro-economic policy and long-term contracts, are complements rather than strict alternatives. Indeed, when extensive empirical tests of these and other hypotheses are carried out, a range of explanations will surely be necessary to account for the important changes in cycle behavior that are documented in Sections I and II of this paper. Given the crucial importance of the Phillips curve slope for macro-economic policy, as suggested at the end of the last section, this research should soon be undertaken.

[13] The classification of union labor agreements by contract duration yields the following percentage breakdown for 1948 and 1972:

Duration in Years	1948	1972
1	.75	.02
1–2	.10	.15
2^+–3		.11
3;3^+	.15	.57

The 1948 data are from Wladimir Woytinsky and the 1972 data are from *Characteristics of Agreements Covering 1,000 Workers or More.*

[14] Union membership in the manufacturing sector, 1910, is from Leo Wolman. Data for 1973 are found in Richard Freeman and James Medoff (p. 44). The economy-wide percentage union membership is calculated by dividing Series D952 by Series D4 and D14 in *Historical Statistics of the United States.* Note that while duration increased *within* the 1950–76 period, the percentage of the manufacturing labor force covered by collectively bargained agreements declined slightly. According to data in Freeman and Medoff (p. 44), 67 percent were covered in 1958 and 61 percent in 1973–75. Of course, the arguments concerning the spread of coverage apply only to interperiod comparisons.

Appendix—Data Sources

Wage Data

1. Compensation per Hour, Manufacturing—1890–1945: Series B72 of *Long-Term Economic Growth, 1860–1965*. The Rees time-series is expressed in constant 1957 dollars. To get nominal wages, the wage series was multiplied by the *CPI*, Series E 135 in *Long-Term Economic Growth, 1860–1965*. 1950–76: Hourly compensation in manufacturing, found in *BLS* (1978).

Comparisons of wage behavior were made for other series as well. In particular, the results of the paper were consistently verified for:

2. Average Hourly Earnings in Manufacturing—1890–1926: Series D 769-770 of *Historical Statistics of the United States*, developed by Paul Douglas. 1950–1976: Average Hourly Earnings, Manufacturing Sector, *BLS* (1977, p. 81).

The behavior of the economy-wide wages was studied in the current period using Series 745 in *Business Conditions Digest*. This series, "Average Hourly Compensation, All Employees in Private Nonfarm Economy," was found to behave cyclically quite closely with the two recent series in manufacturing given above.

Price Data

1. *GNP* Deflator—1890–1945: Series B62, *Long-Term Economic Growth, 1860–1965*. 1948–76: Series 310, *Business Conditions Digest*.

2. Wholesale Prices—1890–1945: Series B69, *Long-Term Economic Growth, 1869–1965*. 1948–76: Series 330, *Business Conditions Digest*.

Industrial Production—1890–1945: Series A15, *Long-Term Economic Growth, 1860–1965*. 1948–76: Series 47, *Business Conditions Digest*.

Unemployment Rate, All Civilian Workers—1890–1929: Series B1, *Long-Term Economic Growth, 1860–1965*. 1945–76: Series 43, *Business Conditions Digest*.

Money Stock M_2—1890–1929: Series X415, *Historical Statistics of the United States*. 1945–76: Series 102, *Business Conditions Digest*.

REFERENCES

M. N. Baily, "Stabilization Policy and Private Economic Behavior," *Brookings Papers*, Washington 1978, *1*, 11–50.

R. J. Barro, "Unanticipated Money, Output, and the Price Level in the United States," *J. Polit. Econ.*, Aug. 1978, *86*, 549–79.

P. Cagen, "Changes in the Recession Behavior of Wholesale Prices in the 1920's and Post-World War II," *Explor. Econ. Res.*, Winter 1975, *2*, 54–104.

E. Feige and D. Pearce, "Economically Rational Expectations," *J. Polit. Econ.*, June 1976, *84*, 499–522.

R. J. Flanagan, "Wage Interdependence in Unionized Labor Markets," *Brookings Papers*, Washington 1978, *3*, 635–73.

R. Freeman and J. Medoff, "New Estimates of Private Sector Unionism in the United States," Nat. Bur. Econ. Res. disc. paper no. 273, Cambridge, Mass., Aug. 1978.

Milton Friedman and Anna Schwartz, *A Monetary History of the United States, 1867–1960*, Princeton 1963.

Z. Griliches, "Distributed Lags: A Survey," *Econometrica*, Jan. 1967, *35*, 16–49.

H. Gregg Lewis, *Unionism and Relative Wages in the United States*, Chicago 1963.

R. Lucas, "Econometric Testing of the Natural Rate Hypothesis," in Otto Eckstein, ed., *The Econometrics of Price Determination*, Washington 1972.

E. Phelps, "Commodity-Supply Shocks and Full-Employment Monetary Policy," *J. Money, Credit, Banking*, May 1978, *10*, 206–21.

J. Sachs, "A Note on the Phillips Curve and Countercyclical Macroeconomic Policy," unpublished manuscript, 1978.

M. Wachter, "The Changing Cyclical Responsiveness of Wage Inflation," *Brookings Papers*, Washington 1976, *1*, 115–68.

Leo Wolman, *The Growth of American Trade Unions, 1880–1923*, New York 1924.

Wladimir S. Woytinsky, *Employment and Wages in the United States*, Baltimore 1953.

National Bureau of Economic Research, *Annual Report*, New York, various issues.

U.S. Bureau of the Census, *Business Condi-*

tions Digest, Washington, various issues.

______, *Historical Statistics of the United States, Colonial Times to 1970*, Washington 1976.

______, *Long-Term Economic Growth, 1860–1965*, Washington 1966.

U.S. Bureau of Labor Statistics, (*BLS*) *Characteristics of Agreements Covering 1000 Workers or More, July, 1972*, Washington 1972.

______, "Output Per Hour, Hourly Compensation and Unit Labor Costs in Manufacturing, Eleven Countries, 1950–1977," Washington, May 1978.

______, *Handbook of Labor Statistics*, Bull. 1966, Washington 1977.

Family Size and the Distribution of Real Per Capita Income

By EDWARD P. LAZEAR AND ROBERT T. MICHAEL*

Per capita income is an important notion in economics. It is used as an explanatory variable with great frequency in theoretical and empirical analyses, and its size distribution is one of the long-standing topics of economic research. Moreover, it is a concept in which public concern is as deep and sustained as is professional interest. However, information about income is often obtained for household units instead of per capita units, or for only a subset of persons (for example, wage earners). This creates difficult problems with the measurement and, indeed, the concept of per capita income. The problems include: 1) within any household the apportionment of household income to members is not in general known; 2) comparison of household income per capita among households of different structures requires judgment about the relationship between real income and family size. Remarkably little study has been done on the first of these two issues.[1] This paper is another contribution to the vast literature which addresses the second issue.

This latter issue is usually characterized as one of determining the income equivalence among households of various sizes. These equivalence scales can either adjust nominal income in different sized households into a common unit (i.e., into income in husband-wife-two-children equivalents) or adjust the number of household members into a common unit (i.e., into the number of full-time adult equivalents).

Many studies have estimated these equivalence scales since Engel in 1895 first estimated the newborn-baby equivalence among households of various sizes. One of two approaches to the estimation of equivalents has generally been used: 1) a revealed preference approach in which household size/structure variables are included in empirical demand studies and the estimated coefficients on these variables are used to infer equivalence; 2) a judgment of "experts" is relied upon to yield equivalence on the basis of some quasi-objective standard (such as daily nutritional needs) and a cost estimate of these items (food) for each household type is then expanded by some factor to approximate an equivalent income level. Most research favors the first approach (see S. J. Prais and Hendrik Houthakker; A. P. Barten; John Muellbauer) while the official *U.S.* poverty level equivalents are based on the Orshansky equivalence measures derived from a presumed nutritionally adequate economy food plan (see B. S. Mahoney). Our work also uses the first of these two approaches; it differs from many of the other studies not in basic concept but in its empirical strategy. While most studies build family composition effects into a relatively formal structural model of demand and impose considerable restriction in order to obtain an estimable system, we use a reduced-form approach which requires much less of the data.

Professional and popular interest in per capita income is predicated on the assumption that income is an observable, mono-

* University of Chicago and National Bureau of Economic Research (NBER), and Stanford University and NBER, respectively. This research was supported by grants to NBER from the Alfred P. Sloan Foundation and the Lilly Endowment. Sharon R. Scott provided excellent research assistance. We wish to acknowledge with thanks helpful suggestions from Gary S. Becker, Victor R. Fuchs, Kyle Johnson, Victoria Lazear, John H. Pencavel, Sharon R. Scott, and an anonymous referee. This is not an official NBER report as it has not been submitted to the Board of Directors for review.

[1]At a theoretical level Paul Samuelson and Gary Becker consider intrahousehold allocations, but empirically we know of few such studies. The estimates of costs of children by A. M. Henderson or T. J. Espenshade might be considered exceptions.

tonic index of economic well-being. One way to characterize the problem of family size equivalence is to ask about differences by family size in the transformation between income and well-being. We suggest in the model developed below that well-being (or utility) is derived from the service flows obtained from market and nonmarket goods and services; the service flows obtained from any particular bundle of market goods depends on the environment in which they are consumed including the quantities of nonmarket goods and services with which they are used. Nominal income adequately indexes the level of market expenditures and hence the bundle of market goods and services. However, its rate of transformation into service flows differs by family size and family structure, because these affect the environment and the nonmarket goods and services with which the market bundle is used. If in circumstance A, a particular market bundle which costs \$1,000 yields 20 units of service flow while in circumstance B that same bundle yields 30 units of service flow, then in real terms the bundle in circumstance B is equivalent to \$1,500 in units of circumstance A. If we know the different rates of transformation between the market goods and the service flows, we can infer levels of real income equivalence among households of various sizes and structures. In this paper we suggest and implement a way of inferring the differences in these rates of transformation from household spending patterns. We calculate implicit deflators by which nominal income in one family size can be converted to its equivalence in some other family size.

The logic of our empirical strategy is as follows: We take as a numeraire an adult living alone in a single person household. For a husband-wife (two-person) household we observe their actual expenditure on some good (say, clothing) and we independently estimate the expenditure these two persons would have made in total had they lived separately in single person households. We contend that the change in their expenditure on clothing in those two circumstances reflects their response to a change in the price of the service flow from clothing, a price change which resulted from the changed environment in which the clothing is used. The organization of the household, the nonmarket goods and time with which the market goods are used, the scale of activities, etc. differ in the two-person household (circumstance B) from the two one-person households (circumstance A), so the service flow from a given bundle of goods (clothing) differs in these two circumstances. Thus at constant market prices the service flow price changes with the circumstance. From knowledge of uncompensated market-price elasticities and our estimates of changes in expenditures from circumstance A (living in single person household) to circumstance B (living in a two-person, husband-wife household) we can infer what price change the couple acts as if it experienced in going from A to B. We estimate this price change for each of several consumption categories which exhaust total consumption. Combining these price changes into a composite index we have a deflator by which nominal income in two-person, husband-wife families can be converted into real income in single person household (numeraire) units. This same procedure can be used for any other family structure as well. We have used this technique to convert several common family structures—two-person, husband-wife families; three-person, husband-wife-child families; four-and five-person, husband-wife, two and three-children families—into single person equivalents.

Before discussing the model in more detail it is appropriate to discuss why we expect the rate of transformation of dollars into service units to differ by family structure. We suggest three mechanisms:

1) *Family goods*: There are certain public goods within the family or household, goods whose consumption by one member does not diminish their availability to other members. Examples abound: electric light in a room, the beauty of art work on the wall, the security provided by a locked bolt on the door, etc. Here, if \$5 provides the man with a securely locked door and \$5 provides the woman with the same living separately, then together its price is to each

\$2.50, a reduction in the price of the service flow resulting from the change in household size.

2) *Scale economies*: Examples include quantity discounts on larger purchases of perishables; less wastage per unit (for example, if the last teaspoon of milk is thrown out with each purchased carton, the larger the carton the smaller the percentage wasted); reduced excess capacity due to indivisibility (a telephone, TV, shower, refrigerator space, etc. is often idle and the utilization rate can be raised by increases in family size).

3) *Complementarity in the use of goods*: Specialization in household duties can result in greater service flow per dollar spent—if goods are combined with time in a productive way, then the more time per unit of good, the higher the marginal product of the good. Money income may not rise proportionately with family size because additional family members supply less time to the labor market. So the ratio of nonmarket time to the quantity of purchased goods may rise. This in turn should raise the flow of services per dollar spent on goods.

For reasons of family (public) goods, scale economies, and division of labor, we expect the rate of transformation between market purchases and real service flows to vary by family size. While most of the examples given suggest a rising level of services per unit of market purchase as family size increases, there may be offsets as well (such as negative externalities from one person's smoking, longer travel distance to work or play for one or for all family members as a result of living in the collectively optimal location).[2] Neither the logic nor the empirical implementation constrains the direction of effects of family size on the rates of transformation or therefore on real income.

I. The Model

Consider an individual with demand for a particular service flow, S_1. That service flow is obtained by the individual using a purchased market input denoted X_1. As an example, the individual may obtain entertainment services as a flow from his purchased market input, a TV set. The rate of transformation between X_1 and S_1 will in general not be independent of the environment in which the individual interacts with X in securing S. For example, the household size in which the individual resides may affect the rate at which X yields units of S. If there are two household members instead of one, a TV set may yield more or less units of entertainment services to the individual (i.e., more if the two share the TV or discuss the show and less if the second member interferes with the viewing). As command over resources in service flow units (Ss) rather than market goods units (Xs) constitutes a measure of real income, it is appropriate to adjust for these differences in rate of transformation when comparing income levels among households of different sizes.

Suppose the individual's demand for S_1 is

$$S_1 = d(P_{s1}, Y) \tag{1}$$

where P_{s1} is the unit price of S_1 and Y is the individual's nominal money income. Let α_1 be the number of units of S_1 produced per unit of market good X_1:

$$\alpha_1 \equiv S_1/X_1 \tag{2}$$

Note that α_1 is thus the average product of X_1 where X_1 is analogous to capital and S_1 is analogous to output. Nonmarket time (analogous to labor input) is suppressed in our model, but if nonmarket time is used with X_1 in the production of S_1, then an increase in time per unit of X_1 would raise the average product of X_1, implying a higher α_1. Then

$$P_{s1} = P_1/\alpha_1 \tag{3}$$

where P_1 is the unit price of the market good X_1. Thus

$$S_1 = d\left(\frac{P_1}{\alpha_1}, Y\right) \equiv g(P_1, Y) \tag{4}$$

[2]Jacob Mincer has found that joint locational choice tends to reduce the wages earned as both workers trade their own earnings for spouse earnings.

with α_1 fixed to the individual, given his environment. From equations (2) and (4) the derived demand for X_1 would be

$$(5) \qquad X_1 = g(P_1, Y)/\alpha_1 \equiv h(P_1, Y)$$

If for any reason α_1 were changed, the demand for S_1 would change and the derived demand for X_1 might be affected. For example, suppose a change in the individual's household environment altered the ratio of S_1/X_1 from α_1 to, say, $\alpha_1(1+J_1)$. If $J_1>0$, the effective price of a unit of S_1 would fall (see equation (3)), and accordingly its quantity demanded would rise. Equation (4) would be

$$(4') \qquad S_1 = g\left(\frac{P_1}{1+J_1}, Y\right)$$

The derived demand for X_1, however, need not rise: as $X_1 = S_1/\alpha_1(1+J)$,

$$(5') \qquad X_1 = h\left(\frac{P_1}{1+J_1}, Y\right)/(1+J_1)$$

Although the fall in the price of P_{s1} assures a rise in the numerator, the denominator offsets this rise, reflecting the additional units of S_1 obtained per unit of X_1. Only if the demand for S_1 is price elastic will the derived demand for X_1 rise with an increase in J_1.[3]

Taking some particular environment as a numeraire one can use equations (5) and (5′) to infer the value of J_1 which converts dollar values in the second environment into units of the numeraire. The term J reflects the percentage by which the price of the item in service units changes as the environment changes from circumstance A (the numeraire) to circumstance B. The nominal dollars spent on X_1 in circumstance B is the equivalent in real (S_1) terms to $(1+J_1)$ times that expenditure in the units of the numeraire. If, for instance, equation (5) reflects the demand for X_1 for a person living in a household of size 1.0, then equations (5) and (5′) yield estimable relationships from which we can infer the J_1 relevant to a person living in a household of size 2.0.

Notice that the role of leisure (nonmarket time), although suppressed, is not ignored. To the extent leisure is used in conjunction with market goods its effects are captured: α_i is the average product of X_i in the production of S_i, and $(1+J_i)$ reflects the change in that average product when an additional household member is added, so a change in the amount of leisure time used per unit of X_i will be reflected in the J_i. If leisure time produced a distinct service S_l by itself, one might add S_l to the set of items studied. We chose not to take this approach for three reasons. First, it seems intuitively unlikely that leisure produces any substantial amount of service without the use of consumer goods. Indeed many "leisure activities" require considerable market goods and services. But if leisure time is always coupled with goods, then the J_i corresponding to changes in the average product of goods will reflect the impact of any change in leisure time.

Second, since most policy decisions are based on a market goods measure of income rather than a "full-income" concept, it is useful to put our equivalence scales in a form consistent with measured income. Third, since our data do not report hours worked, it is not feasible to obtain information on leisure time in our sample; we are therefore unable to treat leisure time as a distinct service item even if we thought that desirable. In the following paragraphs we set out an explicit set of equations from which we can estimate these Js from survey data on expenditure and income.

[3] $$X = d(P_{si}, Y)/\alpha_1(1+J_1)$$

so $$\frac{\partial X}{\partial J} = \frac{1}{\alpha(1+J)}\frac{\partial d(P_{si}, Y)}{\partial J} - \frac{d}{\alpha(1+J)^2}$$

$$= -\left(\frac{X}{1+J}\right)(\varepsilon_{SP_s}+1)$$

or $$\varepsilon_{XJ} = -\left(\frac{J}{1+J}\right)(\varepsilon_{SP_s}+1)$$

hence $$\varepsilon_{XJ} \gtreqless 0 \text{ as } |\varepsilon_{SP_s}| \gtreqless 1$$

We parameterize equation (5) for an individual m:[4]

$$(6)\quad X_{1m}=a_0+a_1P_1+\ldots+a_nP_n+b_1Y_m$$

where P_i is the price of the ith good and Y_m is the person's nominal income. Similarly, assume that another individual, denoted f, has the same demand function and faces the same market prices,

$$(7)\quad X_{1f}=a_0+a_1P_1+\ldots+a_nP_n+b_1Y_f$$

X_{1m} and X_{1f} refer to the amount of X_1 demanded by the two individuals separately if each lives in a household of size 1.0. Thus, the total amount demanded by the two as single individuals is

$$(8)\quad (X_{1m}+X_{1f})=2a_0+2a_1P_1+\ldots+2a_nP_n+b_1(Y_m+Y_f)$$

If the two individuals were to pool resources and live in a single household of size 2, and if as a result there is an effect on the rate of transformation between X_1 and S_1, the effective price of S_1 would be altered. The price would change from P_1/α_1 to $P_1/\alpha_1(1+J_1)$ as described above and likewise for all other consumption items $S_2\cdots S_n$. Thus the *couple's* demand for X_1 would be (from equation (5)):

$$(9)\quad X_{1mf}=\left[2a_0+2a_1\left(\frac{P_1}{1+J_1}\right)+\ldots+2a_n\left(\frac{P_n}{1+J_n}\right)+b_1(Y_{mf})\right]\Big/(1+J_1)$$

While equation (8) represents the demand for X_1 by these two individuals when they live separately, (9) represents their demand when they live together. Forming the difference, equation (8) minus equation (9):

$$(10)\quad (X_{1m}+X_{1f})-X_{1mf}(1+J_1)$$
$$=2a_1P_1\left(1-\frac{1}{1+J_1}\right)+\ldots+2a_nP_n\left(1-\frac{1}{1+J_n}\right)+b_1(Y_m+Y_f-Y_{mf})$$

which can be written as

$$(11)\quad \Delta PX_1\equiv\frac{P_1(X_{1m}+X_{1f})}{P_1X_{1mf}}=(1+J_1)+2\sum_{i=1}^{n}\eta_{1i}\left(1-\frac{1}{1+J_i}\right)+\eta_{X1,Y}(\Delta Y-1)$$

The left-hand side is the ratio of the expenditure on X_1 the two individuals would make if living alone to the expenditure they make if living as a pair; the η_{1i} are uncompensated own- and cross-price elasticities, $\eta_{X1,Y}$ is the income elasticity, and $\Delta Y=(Y_m+Y_f)/Y_{mf}$. An equation comparable to equation (11) can be set out for each of the n market goods.

Before we consider estimation of equation (11), note that if the $J_1, J_2,\ldots,J_n$ for each market good were calculated we could estimate equivalence between nominal income in the numeraire, circumstance A (living alone) and nominal income in circumstance B (living in a household of size 2). If Y_{mf} is the observed nominal income of the couple and Y^*_{mf} is its real income equivalent in single person household units,

$$(12)\quad Y^*_{mf}=Y_{mf}\left(\sum_{i=1}^{n}(1+J_i)w_i\right)\equiv Y_{mf}(1+J)$$

where the w_i is the expenditure weight of item i in the couple's consumption bundle, and J (unsubscripted) is the weighted average of the J_i. (The choice of base for the weights introduces the classic index number problem.) If, for example, the couple experienced a 5 percent increase in the flow of S_i from X_i for all i then $J_i=J=.05$, and a nominal income of say \$5,000 for the couple

[4]Equation (6) should be thought of as an approximation of the true demand curve. As such, it does not have the normal Slutzky properties (except at the mean). We believe, however, that the linear form is much more robust and much less affected by the significant errors-in-variables problem that plagues analyses of this type. In empirical implementation, it behaved more reasonably than the *log* form often used in these studies.

would be the equivalent of $5,000 (1.05)= $5,250 income for the couple in real (S) units of single person income.

The system of n equations of which equation (11) is representative has for each household several variables or parameters: P_iX_{imf}; $P_i(X_{im}+X_{if})$; ΔY; $\eta_{Xi,Y}$; η_{ii};J_i. With known values for the first five of these sets of variables, we can solve for the J_i. The first set (of which there are n elements) is simply the actual expenditure on X_i by the couple. It can be obtained from survey data for persons living in households of size 2.0. The second set (of which there are n elements) reflects the total expenditure on each item by the two persons if they lived separately in households of size 1.0. That counterfactual expenditure can be estimated from survey data on like individuals living in households of sizes 1 and 2 (as described in detail below). The third variable includes the couple's actual nominal income Y_{mf} and the counterfactual income the two would have received if they were living as two separate individuals. The former is available in survey data and the latter is estimable from survey data on like individuals living in households of sizes 1 and 2 (also described below). The fourth and fifth set of variables include the n income elasticities and n own-price elasticities and the $n(n-1)$ cross-price elasticities of demand. These should be available in the economic literature on empirical demand systems. With these five sets of variables known, the system of equations reduces to n equations in n unknowns—the n values of J_i. The system can be used to estimate the J_is which reflect the price changes couples act as if they experience in going from households of size 1 to households of size 2.

In principle, equation (11) could be estimated for each two person husband-wife couple separately, but as the values of $P_i(X_{im}+X_{if})$ and (Y_m+Y_f) are estimated by regression and subject to nonnegligent estimation error, we have chosen instead to use a measure of the average household values of each as a more reliable estimate.[5]

In particular, after estimating $P_i(X_{im}+X_{if})$ for each couple in our data set based on their characteristics, we form a ratio of that estimate for good i to their actual expenditure P_iX_{imf} and then select the median value of that ratio, ΔPX_i, across all two-person husband-wife households.[6] A similar procedure yields a separate ΔPX for each market good and an estimated ΔY. In addition, an analogous procedure yields independent estimates of ΔPX for all goods and of ΔY for households of other sizes and structures as well. These procedures and the estimated price changes and income equivalents are discussed in the following section.[7] To reiterate, these values are used in equation (11) to permit us to solve for the J_i.

II. Empirical Implementation

The data set used in our study is the 1960–61 *BLS* Consumer Expenditure Survey of 13 thousand households. Six expenditure groups are used: food; clothing; trans-

[5]There are two methods of estimating J for the two-person families: use each of the k couples' ΔPX to obtain an estimate of J_{ik} and then average across the k couples to obtain J_i; use the median ΔPX_i and ΔY for the k couples and then estimate J_i from that median. We used the second method. The first has several problems: the value of J is quite sensitive to ΔPX and in some cases J will not be a real number. So measurement error is much more likely to affect each J_{ik} computed separately than it is if J_i is computed from the relatively robust estimate of the median ΔPX.

[6]The median rather than the mean of the estimated ratios is used since this ratio has in its denominator a stochastic variable assumed to be distributed normally, and thus the ratio has a Cauchy distribution for which the moments do not exist. The expectation of estimators obtained using, say, the mean will not exist. As is standard in such cases we assume that the median of the error is zero across all observations and so minimizing the sum of absolute errors is accomplished by use of the median value of the ratio.

[7]Barten proposed a similar scheme. He suggests $U=u(x_1,\ldots,x_n)$ where $x_i=q_i/m_i$ with q the quantity of the purchased good, and $m_i=m_i(b_1,\ldots,b_f)$ where b is the number of family members of a given type and m_i an index of the composition of the family, and shows that "a change in the composition of the family can be translated into terms of a pseudo-price change" (p. 282). His structure is obviously similar to ours, but the suggested research strategy differs. Barten argues that cross-sectional data can be used to estimate price elasticities: using differences in m_i among households as analogues of differences in market prices, differences in spending patterns can yield estimates of "price" elasticities. We suggest, instead, using independent estimates of price elasticities and the actual (estimated) differences in expenditures to infer the changes in prices.

TABLE 1—ESTIMATES OF EXPENDITURE SHIFTS, IMPLIED PRICE CHANGES, AND REAL INCOME DEFLATORS FOR FAMILIES OF SIZE 2–5; FROM 1960–61 *BLS, CES*

Expenditure Item	2-Person Families[a]		3-Person Families[b]		4-Person Families[c]		5-Person Families[d]	
	ΔPX_i	J_i	ΔPX_i	J_i	ΔPX_i	J_i	ΔPX_i	J_i
Food	1.35	0.995	1.65	1.397	1.76	1.665	1.88	1.835
Clothing	1.31	0.848	1.97	1.676	2.20	2.083	2.49	2.431
Transportation	1.16	0.843	1.78	1.668	1.99	2.108	2.26	2.457
Shelter	1.33	0.873	1.20	0.708	1.31	1.040	1.34	1.118
Goods	1.09	0.944	1.63	1.695	1.99	2.366	2.33	2.684
Service	1.12	0.677	1.19	0.800	1.21	1.041	1.34	1.253
J: Weighted Average (group specific wts.)		0.886		1.338		1.728		1.961
Per capita real income equivalent of $10,000 nominal family income $(Y(1+J_j)/j)$		$9,428		$7,795		$6,819		$5,921
Marginal Person $j/(1+J_j)-(j-1)/(1+J_{j-1})$		.061		.222		.184		.222
n	2,918		1,162		1,598		1,061	

[a]Husband and wife.
[b]Husband, wife, and child.
[c]Husband, wife, and two children.
[d]Husband, wife, and three children.

portation; shelter; other goods; other services. The six groups exhaust total current consumption expenditure.[8] For households of size 1.0 (sample size 598) a reduced-form expenditure demand equation is estimated for each of these six items, using as explanatory variables sex, year (1960 or 1961), region, city size, race, age, and education. This equation is used to estimate for larger-sized households the expenditures each family member would have made had he or she lived separately in a household of size 1.0.

Consider the two-person, husband-wife families. We estimate, using these estimated demand equations and the husband's characteristics (education, age, race, city size, region...), the yearly expenditure the husband would have made on food, clothing, etc. had he lived alone. Likewise using that equation and these characteristics we estimate the yearly expenditure the wife would have made if she lived alone.[9] The sum of these estimated expenditures which he and she would have made if each lived separately constitutes our estimate of $P_i(X_{im} + X_{if})$ for item i. That estimate is divided by the couple's actual yearly expenditure on i, P_iX_{imf}, and that ratio ΔPX_i for item $i = 1,\ldots,6$ is calculated for each of the 2,918 two-person husband-wife families. The median value of the distribution of each of those ratios is used as the value of ΔPX_i in the left-hand side of equation (11) for estimating the J_i for two-person households. The estimated values of ΔPX_i are shown in the first column of Table 1. All exceed unity, implying that the predicted expenditure by the couple living separately as two households exceeds the couple's actual expenditure (by amounts ranging from 9 percent for other "goods" to 35 percent for "food" which includes restaurant expenses).

To obtain an estimate of ΔY a similar procedure was employed. We estimated from the single men and women in our sample separate income functions based on the individual's personal characteristics (year, schooling level, age, race, city size, and region). We then used these equations to estimate for the husband and for the wife

[8]The Appendix contains details of the data set and estimators described in the text.

[9]The characteristics for the wives in this data set are not generally available so we had to use her husband's age, race, and education.

separately in each of our two-person families the income each pair might have received had he and she remained single (behaving as singles do in terms of labor supply and nonwage income generation). With that estimate of $Y_m + Y_f$ and the couple's actual income Y_{mf}, we formed the ratio ΔY_k for each of the $k = 2{,}918$ couples and determined the median value ΔY. Its value was 1.2362.

Equation (11) also requires uncompensated price elasticities of the six market goods. One would think that the vast literature on demand systems in the past two decades would have produced a consensus about their magnitudes under various conditions. We have not found that consensus and have chosen to use elasticity estimates derived from Michael Abbott and Orley Ashenfelter's study. We selected the set of elasticity estimates from the Stone-Geary linear expenditure system, evaluated at 1960 prices. These elasticities are shown in Table 2.[10] Ideally, we require elasticity estimates derived from household-size-specific expenditure behavior, not estimates derived from observations across households of various sizes. One justification for using the Abbott and Ashenfelter estimates is that average household size changed little over the time span covered by their time-series study, by less than 1.0 person over the entire 38-year period and by less than 0.1 person from 1950 to the end of their time-series, 1967. So we feel these estimates are acceptably close to the conceptually appropriate elasticities.

Given these price and income elasticities and the estimates of ΔPX_i and ΔY, equation (11) can be written for each of the six consumption items yielding a system of six equations in six unknowns, $J_1, \ldots, J_6$. As a set of quadratic equations there are two roots for each J_i and it can be shown that each pair contains a positive and a negative root.[11]

Economic theory tells us which of the two roots is relevant: from equations (2) and (3) we know $P_{si}S_i = P_i X_i$ for any α_i, so if $\Delta PX_i > 1.0$, implying expenditure on X_i is lower in the two-person household than in the two single person households, we know their expenditure in terms of S is also lower. If the price elasticity $|\eta_{ii}| < 1.0$, we know that price and expenditure move in the same direction

[10]The Abbott-Ashenfelter system estimated coefficients for food, clothing, shelter, and other services which we used directly. Our other two items are composites, and we simply took appropriately weighted averages of the separate elasticity estimates. Our transportation item is composed of 37.6 percent auto purchases and 62.4 percent auto operations and public transportation, so we used a weighted average of "durables" (which contained auto purchases) and "transportation services." Likewise, our "goods" item contains, for couples, 42.8 percent house furnishings and equipment and 57.2 percent tobacco, recreational expenses, reading material, and motels, so we used a weighted average of "durables" and "other nondurables" (tobacco, oil and gas, other miscellaneous nondurables). The Abbott-Ashenfelter system also includes a demand curve for leisure time. We adjusted the discretionary income slope coefficients on the expenditure items to remove discretionary leisure from the demand system.

[11]Equation (11) for item 1 can be written as

$$(J_1)^2 + \left[2 + 2\eta_{11} - \Delta PX_1 + \eta_{X1,Y}(\Delta Y - 1) + \sum_{i=2}^{n}\left(2\eta_{1i}\frac{J_i}{1+J_i}\right)\right]J_1$$

$$+ \left[1 - \Delta PX_1 + \eta_{X1,Y}(\Delta Y - 1) + \sum_{i=2}^{n}\left(2\eta_{1i}\frac{J_i}{1+J_i}\right)\right] = 0$$

a quadratic equation with two roots for J_1 for given values of $J_2, \ldots, J_6$. Solving the six-equation set of quadratic equations simultaneously yields the consistent set of two roots for each of the six J's. From the quadratic equation with roots

$$r_1, r_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

we know $(r_1)(r_2) = c/a$. In the above quadratic equation $a = 1$, and c equals the second, long term in brackets. For practically every item (k) in our study the sign of $1 - \Delta PX_k + \eta_{Xk,Y}(\Delta Y - 1) < 0$ and since $\eta_{ki} < 0$ and we know $J_i > 0$ the final expression in c is also negative—hence $c < 0$ and thus $c/a < 0$ implying the product $(r_1)(r_2) < 0$ which implies that one root must be negative and one positive. If each of the six Js has one positive and one negative root, and we know from economic theory that a positive J is necessary in light of our estimated ΔPX_i and η_{ii}, then only one set of the many possible combinations of roots to the six-equation system is relevant: the one set of six positive roots. (The one remaining point to be made is that for those few cases in our study for which $1 - \Delta PX_k + \eta_{Xk,Y}(\Delta Y - 1) > 0$ the final expression is many-fold larger in absolute value and negative in sign, hence for these cases as well $c < 0$ and our proof holds.)

TABLE 2—ESTIMATED PRICE AND INCOME ELASTICITIES η_{ij} FOR SIX MARKET GOODS EVALUATED AT 1960 PRICES

i	j = Food	Clothing	Transport	Shelter	Goods	Services	Income
Food	−.631	−.034	−.013	−.085	−.030	−.057	.766
Clothing	−.093	−.507	−.014	−.092	−.032	−.062	.828
Transport	−.124	−.049	−.598	−.122	−.048	−.083	1.101
Shelter	−.100	−.040	−.015	−.555	−.035	−.066	.886
Goods	−.149	−.059	−.023	−.146	−.748	−.099	1.318
Services	−.112	−.044	−.016	−.110	−.039	−.602	.994

Source: Derived from Abbott and Ashenfelter estimates from augmented Stone-Geary linear expenditure system.

along the demand curve. Thus, if $|\eta_{ii}|<1.0$ and $\Delta PX_i>1.0$, we can infer that the price (P_{si}) the couple faces is *lower* than the price the two as single individuals face: that is, the price *fell* from P_{Xi}/α_i to $P_{Xi}/\alpha_i(1+J_i)$. Hence we know

$$\frac{P_{Xi}}{\alpha_i}>\frac{P_{Xi}}{\alpha_i(1+J_i)}\rightarrow(1+J_i)>1\rightarrow J_i>0$$

If instead $|\eta_{ii}|>1.0$, the same logic implies that if $\Delta PX_i>1.0$, then $J_i<0$. Or, of course, if the expenditure rose, $\Delta PX_i<1.0$, then if $|\eta_{ii}|<1.0$, $J_i<0$. The four logical possibilities for the sign of J_i are:

Uncompensated Price Elasticity $\lvert\eta_{ii}\rvert$	ΔPX_i >1.0	<1.0[12]
<1.0	positive	negative
>1.0	negative	positive

As all the own-price elasticities estimated from Abbott and Ashenfelter are inelastic, and all the ΔPX_i shown in the first column of Table 1 happen to be greater than 1.0, all the relevant Js are positive, in this case.[13]

The second column of Table 1 shows the implied value of J_i for two-person households. Weighting by the average expenditure shares of these six items for the 2,918 couples, the average J is also shown. As $J=0.886$, on average the prices of S faced by couples are estimated to be only 53 percent $(1/(1+J))$ as high as the prices of S faced by single person households. The relative prices also are affected: the price of market "services" falls least (to 60 percent of its single person level $(1/1+0.677)$), while the price of services from "food" falls most (to 50 percent of its single personal level $1/(1+0.995)$). The largest gains from the change in circumstance (marrying, sharing duties, achieving economies of scale and joint consumption advantages) are in food, goods and shelter; the smallest gains are in the purchase of market services, comprised primarily of medical care and personal care expenditures. We view this set of estimated price changes as intuitively plausible. Its overall magnitude is large and that we discuss below.

Using equation (12) it is a simple matter to convert nominal family income of a couple into its single person equivalent for the two individuals: if the couple's observed nominal family income is \$10,000, the "real" income is \$10,000 $(1+J)$ = \$18,856, as per capita, the husband and wife realize \$18,856 ÷ 2 = \$9,428, as indicated in Table 1. That is, a couple with observed family income of \$10,000 is estimated to have the same *per capita* income as a single person with \$9,428. The "economies of scale" are substantial, although these gains are not simply scale effects but also include the effects of complementarity in time and money use and the public (family) goods effects. As mentioned earlier, we can convert dollars into common units per person or convert persons into full-time full-person equivalents, and doing so in this case yields 1.061 as the full-person

[12] Real roots may not exist.

[13] It may be well to stress that neither of these conditions (inelasticity nor decline in expenditure) is a logical or computational necessity. In fact, in preliminary work we used other estimates of elasticities, one of which exceeded 1.0, and in other preliminary work we did estimate a few ΔPX_i's which were <1.0. A different level of commodity aggregation would surely yield such estimates (for example, smaller aggregates will have more substitutes so more elastic demand curves). The procedure used here can accommodate these differences easily.

TABLE 3—NOMINAL AND REAL FAMILY INCOME, Y AND Y^*, BY FAMILY SIZE (MEAN (STANDARD DEVIATION) AND MEDIAN)

	Family size					
	1	2	3	4	5	Total
			Mean (standard deviation)			
Family income						
Nominal (Y)	$3539.	$5099.	$6141.	$6990.	$7221.	$5856.
	(2353.)	(3955.)	(3135.)	(4143.)	(4711.)	(4049.)
Real (Y^*)	3539.	9611.	14382.	18903.	21203.	13572.
	(2353.)	(7439.)	(7508.)	(9858.)	(12330.)	(10181.)
Per capita income						
Nominal (Y)	3539.	2550.	2047.	1748.	1444.	2216.
	(2353.)	(1977.)	(1045.)	(1036.)	(942.)	(1693.)
Real (Y^*)	3539.	4806.	4794.	4726.	4241.	4601.
	(2353.)	(3719.)	(2503.)	(2465.)	(2466.)	(3046.)
			Median			
Family income						
Nominal (Y)	3324.	4369.	5566.	6381.	6576.	5326.
Real (Y^*)	3324.	8238.	12974.	17393.	19438.	11987.
Per capita income						
Nominal (Y)	3324.	2184.	1855.	1595.	1315.	1805.
Real (Y^*)	3324.	4119.	4324.	4348.	3888.	4125.
n	598	2918	1162	1598	1061	7337

equivalent size of a husband-wife family, $2/(1+J)$. Here again the single person is taken as the numeraire.

To obtain estimates of equivalents for families of size 3.0 we proceed in the same manner. We chose to use only three-person households comprised of a husband-wife and child under age 18; the survey had 1,162 such families. For each person in the family we again used the demand equations estimated on single individuals to estimate what each family member would have spent on each of the six consumption items had he or she lived in a single person household.[14] Summing these three estimated expenditures together gave us the estimated expenditure used in the numerator of equation (11), and the family's actual expenditure is used in the denominator to calculate the ΔPX_i for each item for each of the 1,162 husband-wife and child families. The median value of the distribution for each item is shown in the third column in Table 1. For each family we again estimated a ΔY and its median value for these 1,162 families was 1.2699. Using the income and uncompensated price elasticities shown in Table 2, the J's shown in the fourth column of Table 1 were calculated. Again, all the prices are estimated to have fallen, with the decline for shelter and services relatively low and the biggest gains appearing to be in goods, transportation, and clothing. Here the average J is 1.3384, so the prices faced by these three-person families is estimated to be, on average, only 43 percent as high as the prices faced by the three single person households.

The same procedure has also been applied to the 1,598 four-person (husband-wife and two children) families and the results are shown in Table 1. Finally, the procedure was applied to 1,061 five-person (husband-wife and three children) families as well, and these results are also shown in Table 1. (The median values of $\Delta Y(4)$ and $\Delta Y(5)$ are 1.17 and 1.16, respectively.)

Throughout, our estimates imply quite large gains in real terms from increases in family size. Real income (income in service flow equivalents with single persons as the

[14]For children in the family we knew age brackets: 0-6, 6-12, 12-17 (and for larger sized families > 18). So we used piece-wise linear demand equations and assigned ages 3, 9, 15 (and 21) and assigned education levels 0, 3, 9 (and 15), respectively. For the other characteristics of the child, we used father's characteristics (for race, city size, region, year) and for sex we used the value 0.5.

numeraire) is substantially higher in large families than is nominal income. We estimate that $10,000 nominal income to a family of five yields each of the five members the equivalent of $5,922 in a single person equivalent dollars, or said differently, five can live together about as cheaply as 2.0 can live separately.

While Table 1 shows estimates of J and equivalents by family size, we have taken a few steps toward estimating household-specific Js. We computed expenditure weights for specific family types defined over four age-of-head and five nominal-income groups and, using the J_i estimates in Table 1, estimated a J for each of the twenty family types. For smoothing purposes, for each family size we ran a separate weighted regression across these twenty Js and used those regressions to assign a J to each household based on its size, age of head, and nominal income.

As our initial intention was to obtain comparable per capita income measures across families of different sizes, we can use these Js to do so. For each of the five family types discussed above (single persons,..., husband-wife and three children families), we assign a J ($J=0$ for the single persons, the numeraire).[15] There are 7,337 such households in the *BLS* data for which we then have two measures of family income:

Y_j: the jth household's actual nominal family income;

$Y_j^* = Y_j(1+J_j)$: the jth household's real income equivalent.

Table 3 shows the mean and median Y, Y^*, and per capita Y and Y^* by family size. In single person equivalent dollars the real family income of larger families is quite high, but the real per capita income is comparatively constant across families of different sizes. As a result, the overall (median) per capita *real* income is substantially higher than the per capita *nominal* income among families of sizes 1 to 5 ($4,125 compared to $1,805). Figure 1 shows the frequency distribution of per capita Y and Y^*, emphasizing the far more evenly distributed and larger mean value of the latter.

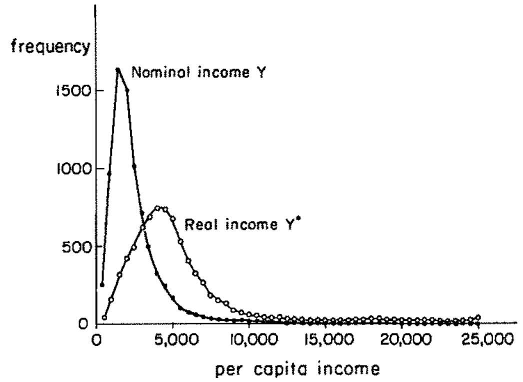


FIGURE 1. DISTRIBUTIONS OF PER CAPITA NOMINAL INCOME Y AND PER CAPITA REAL INCOME Y^* FOR 7,337 HOUSEHOLDS OF SIZE 1–5

III. Interpretation

Our estimates suggest that a substantial adjustment in nominal income is necessary to reflect real (single person equivalent) income among families of different sizes. Scale economies, joint consumption of goods, and complementarity of goods and nonmarket time account for these substantial adjustments. Other studies have also emphasized these factors separately, as, for example, I. A. Sirageldin's estimates of the distribution of real income including the market value of the household tasks performed by women, or Reuben Gronau's estimates of the housewife's contribution to full income.

We find that the adjustment in median income in our sample of families and unrelated individuals raises a median nominal per capita income of $1,805 to $4,125 in real dollars. One might ask what portion of that adjustment is capturing the nonworking wife's contribution to full income. Gronau's estimates (using information on time use from the Michigan Income Dynamics data) suggest that for married women of all ages and education levels combined, the nonmarket work done by women would raise income by about 60 percent (see his Table 7). Our numbers suggest a far greater adjust-

[15]We have implicitly assumed homogeneity among single persons. We might have instead estimated the prices faced by single persons of various ages or sex and converted all singles into, say, 40-year-old male equivalents.

TABLE 4—EQUIVALENCE SCALES FOR HOUSEHOLDS OF SIZE 1–5, ESTIMATED BY FOUR TECHNIQUES

Household size	Naive	*BLS*	Orshansky	LM
Panel A Equivalence: 4 Person = 100				
1 Person (male)	25	36	53	68
2 Person (husband-wife)	50	60	67	72
3 Person (husband-wife-child < 18)	75	82	80	88
4 Person (husband-wife-2 children)	100	100	100	100
5 Person (husband-wife-3 children)	125	116	118	115
Panel B Equivalence: 1 Person = 100				
1 Person (male)	100	100	100	100
2 Person (husband-wife)	200	167	126	106
3 Person (husband-wife-child < 18)	300	228	151	128
4 Person (husband-wife-2 children)	400	278	189	147
5 Person (husband-wife-3 children)	500	322	223	169

ment, but ours reflect not only this nonmarket time effect but also scale and joint consumption effects. Our estimate for each family of $(Y_m + Y_f)$, discussed above and used in estimating ΔY for each household size separately, indicates the nominal income the family would have received if both spouses allocated their time as comparable single persons do. Thus we can consider $Y' = Y'_m + Y'_f$ as a crude measure of the labor supply-adjusted family income. If we compare its per capita value with the per capita observed nominal income (Y) and our estimate of the per capita real income (Y^*), we find:

	Per Capita income	
	Median	Mean
Nominal income (Y):	\$1,805	\$2,216
Labor-supply adjusted income (Y'):	2,270	2,480
Real income (Y^*):	4,125	4,601

The adjustment for labor supply accounts for only a small portion of the adjustment from nominal to real income. The remainder may be scale effects, joint consumption effects, possibly complementary effects of one spouse on the other, or other effects. It does not appear to be the case that most of our estimated adjustment of Y to Y^* is simply an adjustment for unearned income of married women.

Another way to isolate the effects of nonmarket time is to look separately at the implied price effect (or J) for families with two employed adults and those with only one employed adult. We would expect a higher J in the families with only one employed adult, reflecting the greater flow of services (S) per dollar of market goods (X) as discussed above. We selected, from our sample of 2,918 two-person families, two subsets: (a) a group of 1,043 comprised of all those couples for which the head was employed full time and the spouse was not employed during the year; (b) a group of 362 comprised of all those couples for which both the head and the spouse were employed full time. On the basis of the observed ΔPX_i and ΔY for these two subgroups separately, the J for the one-earner families was 0.83, the J for the two-full-time-earner families was 0.75. While the difference is in the expected direction, the J for the two-earner families is higher than might be expected. This suggestive evidence corroborates the conclusion of the previous

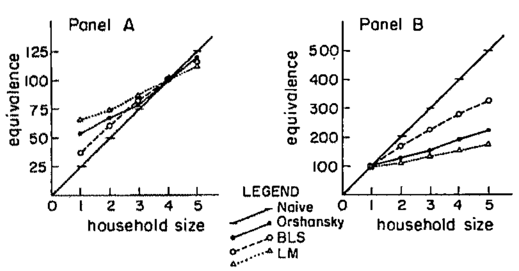


FIGURE 2. EQUIVALENCE SCALES FOR HOUSEHOLDS OF SIZE 1–5 BY FOUR TECHNIQUES

TABLE 5—ESTIMATED PERCENTAGE IN POVERTY
(1960–1961 SAMPLE; IN PERCENTAGE)

	Age of head			
	<35	35–64	>65	Total
Orshansky scale:	6.7	8.9	25.0	11.0
LM scale:	7.7	10.1	33.7	13.4
n	(1762)	(4365)	(1210)	(7337)

	Family size					
	1	2	3	4	5	Total
Orshansky scale:	17.6	14.3	6.9	6.7	9.3	11.0
LM scale:	25.3	18.4	8.2	6.9	8.7	13.4
n	(598)	(2918)	(1162)	(1598)	(1061)	(7337)

paragraph that much of the gain in real income comes from sources other than the differences in labor supply.

Another real income adjustment found frequently in the literature is the equivalence scale for households of different sizes and age structures. Perhaps the two best known equivalence scales are the *BLS* and the Orshansky scales. The *BLS* scale is used in their Family Budget series and derived by inferences based on the notion that families who spend the same proportion of their disposable income on an income-inelastic item such as food act as if they have the same real income. The Orshansky scale is used in constructing poverty levels for families of various sizes, and derived from estimates of the costs of purchasing nutritionally adequate diets for families of different sizes. As our Table 1 indicates the per capita single person equivalent income of \$10,000 in nominal family income, we can use these figures to derive a comparable equivalence scale. That is, $Y_0((1+J_j)/j)/((1+J_k)/k)$ would be the equivalent nominal income in k-person household units of the nominal income of Y_0 in j-person households.[16] Table 4 shows these equivalence scales for *BLS*, Orshansky, and LM (Lazear-Michael) estimates. For comparison the first column shows the numbers for the naive assumption of complete absence of scale or other effects. Panel *A* indexes these equivalence scales based on a family of size four and since the choice of a numeraire can affect the apparent differences, Panel *B* shows the indices based on a single person equivalence. Also see Figure 2.

Given the tremendously different algorithms used in constructing these three equivalence scales, it is interesting to note how similar they are, at least among families of sizes 2 through 5. Large differences exist, however, in going to single person households, where the LM estimate suggests far more "scale" economy than the other (especially the *BLS*) estimates. This difference for single person households is especially important for comparisons of equivalent real income over time, as the proportion of single person households has risen considerably in the *U.S.* population in the postwar period.

One other comparison which can be performed with our data is to identify the differences in the poverty population when that population is designated by the official equivalence scales and by the LM equivalence scales. We calculated the two poverty benchmarks for each of the 7,337 households in our sample and determined which households were "in poverty" by each definition. The official level of poverty income for a family of four composed of husband, wife, and two children in 1960 was \$3,022 and in 1961, \$3,054. Using the Orshansky equivalence scales in Panel *A* of Table 4 the poverty level for the other types of households are easily computed (for example, for

[16]We convert nominal income Y_0 in a j-person household into real per capita income in single person household units as $Y_0(1+J_j)/j \equiv Y_j^*$. This is converted to per capita income units of a k-person household as $(Y_j^*/(1+J_k))$. Or to convert to family income for that household, multiply the expression by k to obtain $(Y_j^*/(1+J_k))k$.

a single person under age 65 in 1960, the poverty level was (.53)($3,022.) = $1,602). Likewise, using the LM estimates of equivalence the poverty level comparable to $3,022 for a family of four is easily computed (for example, for a single person under 65 in 1960, (.68)($3,022.) = $2,055).

The percentages in poverty, by age of head, and by family size, using the Orshansky scale and the LM scale are shown in Table 5. While the overall percentage is only moderately higher with the LM equivalence scales, a substantially higher fraction of single persons or older persons is estimated as "in poverty" using the LM scales. Given the substantial "scale" effects between single person and four-person households, if $3,000 is the benchmark for poverty for the four-person family the single person requires a relatively larger amount to be as well off.

There are other comparisons and modifications which might be made. We hope in subsequent work to use more recent data, containing more complete information on each family member's demographic characteristics. Our procedure could easily be extended to families of sizes greater than five and to different family structures (for example, female-headed families). Relaxation of the assumed homogeneity of all single person households in terms of Js seems an appropriate extension as does some additional checking on the sensitivity of our estimates to the price elasticities which were employed here.[17] We suggest our procedure for estimating per capita income equivalence among families of different sizes has among its other appealing properties the fact that it is embedded in a standard economic theory of demand. We use changes in expenditures plus price elasticities to infer changes in prices from which real (price deflated) levels of income can be inferred.

TABLE A1—DETAILS ABOUT THE SIX EXPENDITURE CATEGORIES; 1960–61 *BLS, CES*
(Weights for two-person families)

	Category Percent	Total Percent	Dollars
Food	100.0	25.9	1079.
Food at home	75.5		
Food away	16.9		
Alcohol	7.6		
Clothing	100.0	8.4	350.
Transportation	100.0	16.8	699.
Auto operation	51.0		
Car purchase	37.6		
Public transportation	11.4		
Goods	100.0	13.0	540.
Household furnishings and equipment	42.8		
Recreation	27.2		
Reading	7.8		
Vacation homes	1.3		
Lodging away from home	5.1		
Tobacco	15.8		
Services	100.0	11.3	469.
Medical care	71.5		
Personal care	26.4		
Education	2.1		
Shelter	100.0	24.6	1022.
Utilities	23.3		
Household operation	24.3		
Rent or owner's expenses	52.4		

[17]Substantial sensitivity analysis has been done. In addition to the set of estimates described in this paper, three other sets of estimates have been made. Elasticity estimates from the Abbott-Ashenfelter system, augmented to include leisure as a separate item, were used directly in one set of estimates (estimate A). Using the own-price elasticities but assuming all cross elasticities to be zero and setting $\Delta Y = 0$, a second set of J_i's were computed (estimate B). Similarly, using a somewhat different set of own- and cross-price elasticities (for example, the elasticity of services with respect to the price of food $= -0.095$ in estimate A but $= -0.149$ in estimate C) and a slightly different set of $\eta_{Xi,Y}$, another set of J_i's were estimated (estimate C). The estimated J for families of size 2 through 5 from these three sets of estimates were

Text estimates: .89; 1.34; 1.73; 1.96
Estimate A: .84; 1.30; 1.67; 1.92
Estimate B: .76; 1.23; 1.45; 1.68
Estimate C: .83; 1.26; 1.63; 1.89

So while there are not inconsequential differences and the estimates reported in the text are somewhat larger than the others, rather large differences in the initial elasticities yield only modestly different average Js. We think this comparison implies a substantial degree of robustness in our estimates.

TABLE A2—REDUCED-FORM EXPENDITURE EQUATIONS; SINGLE CONSUMERS; 1960–61 *BLS, CES*

Variable	Food	Clothing	Transport	Shelter	Goods	Services	Sample Mean
Intercept	149.84	308.000	−204.05	−229.78	78.23	66.92	1.00
	(0.57)	(2.98)	(−0.57)	(−0.97)	(0.48)	(0.55)	
Sex	394.46	−96.31	195.02	−21.41	126.47	−77.48	0.46
1 = male	(8.18)	(−5.11)	(3.01)	(−0.50)	(4.27)	(−3.51)	(0.50)
Year	−98.99	24.00	−7.42	17.64	27.93	21.04	0.57
1 = 1961	(−2.14)	(1.33)	(−0.12)	(0.42)	(0.98)	(0.99)	(0.50)
North East	35.83	0.02	12.06	72.57	106.32	−65.05	0.28
	(0.53)	(0.01)	(0.13)	(1.21)	(2.54)	(−2.08)	(0.45)
N. Central	−138.15	−67.03	−153.98	−117.47	25.80	−85.03	0.29
	(−2.04)	(−2.52)	(−1.69)	(−1.93)	(0.62)	(−2.74)	(0.45)
South	−192.98	10.83	−92.61	−119.33	55.76	−69.86	0.24
	(−2.76)	(0.40)	(−0.99)	(−1.93)	(1.30)	(−2.18)	(0.43)
Central City	133.62	21.77	−110.87	179.03	102.12	60.03	0.48
	(2.74)	(1.14)	(−1.70)	(4.04)	(3.41)	(2.69)	(0.50)
Large City	63.84	43.20	172.78	194.20	61.12	−35.49	0.05
	(0.59)	(1.03)	(1.20)	(2.02)	(0.93)	(−0.72)	(0.22)
Race	161.66	48.77	78.45	137.78	106.54	49.56	0.89
1 = white	(2.11)	(1.63)	(0.76)	(2.11)	(2.26)	(1.41)	(0.31)
Age < 25	4.90	−4.86	14.14	6.34	−12.21	−0.06	22.44
	(0.48)	(−1.21)	(1.02)	(0.70)	(−1.93)	(0.01)	(2.08)
Age 25–34	6.51	−3.30	12.24	8.72	−4.99	1.21	29.56
	(0.84)	(−1.09)	(1.17)	(1.26)	(−1.04)	(6.34)	(2.73)
Age 35–54	6.64	−3.79	4.10	9.70	−5.01	0.81	46.22
	(1.36)	(−2.00)	(0.62)	(2.19)	(−1.66)	(0.36)	(6.32)
Age 55–64	1.71	−4.54	0.80	5.68	−5.02	0.79	60.44
	(0.44)	(−3.00)	(0.15)	(1.61)	(−2.10)	(0.44)	(2.49)
Age > 65	−0.46	−4.87	−0.99	5.63	−5.40	0.57	72.43
	(−0.14)	(−3.89)	(−0.23)	(1.95)	(−2.74)	(0.39)	(5.53)
Education	21.10	14.94	39.62	43.52	21.90	15.29	11.29
	(3.34)	(6.05)	(4.67)	(7.40)	(5.65)	(5.29)	(4.10)
R^2	0.22	0.34	0.14	0.25	0.21	0.14	
Mean	799.30	280.15	451.28	779.56	328.62	260.46	
S.D.	(604.63)	(257.52)	(773.89)	(477.53)	(369.85)	(263.46)	
n	598	598	598	435	598	598	

TABLE A3—*OLS* REGRESSIONS ON INCOME, SINGLE CONSUMERS, BY SEX

	Male Income	Female Income	Sample Mean Men	Sample Mean Women
Constant	−924.13	−5477.61	1.00	1.00
	(−0.66)	(−6.48)		
Year	−155.96	423.07	0.60	0.54
1 = 1961	(−0.51)	(2.39)	(0.49)	(0.50)
North East	412.87	−166.24	0.26	0.30
	(0.95)	(−0.61)	(0.44)	(0.46)
North Central	−370.95	−452.18	0.31	0.28
	(−0.89)	(−1.63)	(0.46)	(0.45)
South	−509.85	−295.20	0.22	0.26
	(−1.15)	(−1.06)	(0.41)	(0.44)
Central City	8.72	564.01	0.44	0.52
	(0.03)	(3.09)	(0.50)	(0.50)
Large City	2216.00	376.80	0.05	0.05
	(3.13)	(0.92)	(0.21)	(0.22)
Race	409.78	1152.73	0.85	0.93
1 = white	(0.91)	(3.28)	(0.36)	(0.25)
Age	128.34	210.02	45.00	49.52
	(2.44)	(7.59)	(17.74)	(18.74)
Age^2	−1.54	−2.17	2336.73	2803.07
	(−2.89)	(−7.84)	(1700.59)	(1871.60)
Education	205.82	269.87	10.20	12.21
	(4.89)	(11.00)	(4.34)	(3.64)
R^2	0.26	0.43		
n	275	323		
Mean	3578.70	3505.31		
S.D.	2707.08	1996.81		

TABLE A4—PERCENTAGE OF FAMILIES IN POVERTY (1960–61 SAMPLE), USING ORSHANSKY AND LM EQUIVALENCE SCALES, BY AGE OF HEAD AND FAMILY COMPOSITION

Family Composition	Age of Head <35	35–64	65+	All ages
Orshansky Scale:				
1 Person	4.0	15.7	41.7	17.6
2 Persons	2.8	11.0	22.6	14.3
3 Persons	6.6	6.5	29.4[a]	6.9
with child < 6	7.2	4.2	—	6.6
with child age 6–17	3.2	6.9	29.4[a]	7.1
4 Persons	7.4	6.0	41.7[a]	6.7
all children < 6	7.4	8.6	—	7.6
with children > 6	7.3	5.8	41.7[a]	6.4
5 Persons	10.7	8.4[a]	33.3[a]	9.3
with children < 6	12.2	27.3	—	13.8
all children > 6	10.1	8.1	33.3[a]	8.8
Total	6.7	8.9	25.0	11.0
Lazear-Michael scale:				
1 Person	9.0	22.8	54.6	25.3
2 Persons	3.9	12.7	31.0	18.4
3 Persons	8.0	7.8	29.4[a]	8.2
with child < 6	8.8	4.2	—	7.9
with child 6–17	3.2	8.3	29.4[a]	8.4
4 Persons	7.9	6.0	50.0[a]	6.9
all children < 6	8.0	8.6	—	8.1
with children > 6	7.8	5.8	50.0[a]	6.6
5 Persons	9.5	8.0[a]	33.3[a]	8.7
with children < 6	12.2	27.3	—	13.8
all children > 6	8.4	7.7	33.3[a]	8.1
Total	7.7	10.1	33.7	13.4

[a]Cell contains less than twenty families.

REFERENCES

M. Abbott and O. Ashenfelter, "Labour Supply, Commodity Demand, and the Allocation of Time," *Rev. Econ. Stud.*, Oct. 1976, *43*, 389–411. (Estimation corrected Oct. 1977.)

A. P. Barten, "Family Composition, Prices and Expenditure Patterns," in P. E. Hart et al., eds., *Econometric Analysis for National Economic Planning*, London 1964.

G. S. Becker, "A Theory of Social Interactions," *J. Polit. Econ.*, Dec. 1974, *82*, 1063–93.

E. Engel, "Die Lebenskosten belgischer Arbeiter—Familien Fruher und jetzt," *Int. Statist. Inst. Bull.*, No. 9, 1895.

T. J. Espenshade, "The Cost of Children in Urban U.S.," Population Monog. Series No. 14, Inst. Int. Stud., Univ. California-Berkeley 1973.

R. Gronau, "Home Production—A Forgotten Industry ," *Rev. Econ. Statist.*, forthcoming..

A. M. Henderson, "The Cost of Children, Part I, " *Pop. Stud.*, Sept. 1949, *3*, 130–50; "Parts II, III," Dec. 1950, *4*, 267–98.

B. S. Mahoney, *The Measure of Poverty*, Poverty Studies Task Force, H. E. W., Washington 1976.

J. Mincer, "Family Migration Decisions," *J. Polit. Econ.*, Oct. 1978, *86*, 749–773.

J. Muellbauer, "Testing the Barten Model of Household Composition Effects and the Cost of Children," *Econ. J.*, Sept. 1977, *87*, 460–487.

M. Orshansky, "Table 11," in B. S. Mahoney, ed., *The Measure of Poverty*, Washington 1976.

S. J. Prais and Hendrick S. Houthakker, *The Analysis of Family Budgets*, London 1955.

P. A. Samuelson, "Social Indifference Curves," *Quart. J. Econ.*, Feb. 1956, *70*, 1–22.

I. A. Sirageldin, "Nonmarket Components of National Income," Inst. Soc. Res., Univ. Michigan, Ann Arbor 1969.

U.S. Bureau of Labor Statistics (*BLS*), *Survey of Consumer Expenditures 1960–61*, Washington 1966.

Sources of Quality Change in Labor Input

By PETER CHINLOY*

Labor input is the product of total hours worked and average labor quality per hour. Labor quality accounts for the level of skill provided per hour worked, including educational and demographic factors. Change in labor quality can be expressed as the sum of main effects associated with these factors and interactive effects of various orders yielding a growth accounting equation for labor input. This is applied to a classification of total hours worked by sex, class of worker (employee or self-employed), age, education, and occupation for the *U.S.* private domestic economy 1947–74. The main conclusions are:

(a) The contribution to labor input growth of education is 0.6 percent per annum. This effect is reduced by one-half if interactive effects are included, as the educated become younger and more female. The main effect for education declines by over one-fifth between 1959–63 and 1971–74, which may indicate a decline in the contribution of education to *U.S.* productivity growth.

(b) A linear logarithmic quality change estimate excluding interactions overstates the growth of labor quality by one-half.[1] This suggests that the contribution of education and experience to economic growth, for example, may not be measured by multiplying together indices of each factor. The overstatement amounts to 0.3 percent per annum, which at a labor share of two-thirds, overstates the contribution of labor input to output growth by 0.2 percent per annum.

(c) The main effect of the substantial increase in relative share of women in total hours is negative. The inclusion of interactive effects reduces this effect from $-.15$ to $-.07$ percent per annum, accounting for the skill composition of women.

(d) A watershed develops in the 1959–63 period in assessing the relative importance of total hours and labor quality as sources of labor input growth. Over 1947–59, labor quality is relatively dominant, accounting for over three-quarters of labor input growth of about 1.3 percent per annum. For 1963–74, labor input increases in growth to 1.9 percent annually, but quality change accounts for only one-tenth. Quality change in the *U.S.* labor market almost disappears, declining from 1.12 percent for 1947–52 to 0.12 percent over 1971–74.

I. Labor Input Indexing

The indexing of labor input commences with a production function aggregating nonlabor services and the services provided by different types of labor. An aggregate of labor input exists if types of labor are weakly separable from nonlabor inputs. I assume the labor market is efficient, and types of labor paid marginal products. Labor input can increase even if total hours worked are constant. Suppose there are two types of labor, skilled and unskilled. The former receive above average wages, and both work the same hours. If an unskilled worker becomes skilled, total hours remain unchanged, but labor input increases since the marginal product of this worker increases. The objective is to quantify these changes in labor input and associate them with characteristics of employment.

The production function, separable between labor and nonlabor inputs, is at time t:

$$(1) \qquad y_t = g(z_t, x_{1t}, \ldots, x_{pt}, t)$$

where y_t represents output, z_t labor input,

*University of British Columbia. I am grateful to William Barger, Erwin Diewert, Richard Freeman, Frank Gollop, Robert Hall, and Dale Jorgenson for their substantial contributions to this work.

[1]Richard Nelson has argued that experienced growth may involve interaction effects as well as main effects of factors.

and $x_{1t},\ldots,x_{pt}$ the services of nonlabor inputs.[2] The labor aggregate is

$$(2) \qquad z_t = f(h_{1t},\ldots,h_{It})$$

where $h_{it}, i=1,\ldots,I$ represents hours worked by type i labor. Under efficient labor markets and linear homogeneity of f,

$$(3) \qquad \frac{\partial \ln z_t}{\partial t} = \sum_{i=1}^{I} s_{it} \frac{\partial \ln h_{it}}{\partial t}$$

where

$$(4) \quad s_{it} = \partial \ln f / \partial \ln h_{it} = w_{it} h_{it} \Big/ \sum_{i=1}^{I} w_{it} h_{it}$$

is the share of the ith type in total labor compensation. The hourly wage of the ith type is w_{it}, and its compensation $w_{it}h_{it}$. The growth rate of labor input is a convex combination of growth rates of total hours for each type of labor, with compensation shares as weights.

Total hours worked for all types of labor are $m_t = \Sigma_{i=1}^{I} h_{it}$, and the growth rate of these is

$$(5) \qquad \frac{\partial \ln m_t}{\partial t} = \sum_{i=1}^{I} b_{it} \frac{\partial \ln h_{it}}{\partial t}$$

with $b_{it} = h_{it}/m_t$ the share of total hours worked by the ith labor type. Average labor quality per hour is labor input divided by total hours worked:

$$(6) \qquad a_t = z_t / m_t$$

and its growth rate is

$$(7) \qquad \frac{\partial \ln a_t}{\partial t} = \sum_{i=1}^{I} (s_{it} - b_{it}) \frac{\partial \ln h_{it}}{\partial t}$$

at time t. This growth rate is termed quality change, and is the sum of growth rates of hours worked by each type of labor, weighted by the difference between the shares in labor compensation and hours worked.

The labor aggregate is specified to have the translog form

$$(8) \quad \ln z_t = \alpha_0 + \sum_{i=1}^{I} \alpha_i \ln h_{it} + 1/2 \sum_{i=1}^{I} \sum_{j=1}^{I} \beta_{ij} \ln h_{it} \ln h_{jt}$$

where α_0, α_i, $i=1,\ldots,I$ and β_{ij}, $i,j=1,\ldots,I$ are parameters and where $\beta_{ij}=\beta_{ji}$.[3] Under linear homogeneity,

$$\sum_{i=1}^{I} \alpha_i = 1 \quad \text{and} \quad \sum_{j=1}^{I} \beta_{ij} = 0, \quad i=1,\ldots,I$$

Efficiency requires the relative share of the ith type of labor to be equal to its logarithmic marginal product, or

$$(9) \qquad s_{it} = \alpha_i + \sum_{j=1}^{I} \beta_{ij} \ln h_{jt}$$

where the s_{it} are defined in (4). Equations (8) and (9) plus the symmetry conditions $\beta_{ij}=\beta_{ji}$ imply

$$(10) \qquad d_t = \sum_{i=1}^{I} v_{it} \Delta \ln h_{it}$$

for the logarithmic first difference $d_t \equiv \Delta \ln z_t$, where Δ denotes the first difference operator and $v_{it} = 1/2(s_{it} + s_{i,t+1})$. This is the growth rate of the translog index of labor input.[4]

The growth of hours worked is defined in discrete time by

$$(11) \qquad h_t \equiv \Delta \ln m_t$$

[2]Weak separability of labor and nonlabor inputs, and homotheticity of the direct production function, implies a separable wage aggregate within the dual cost function. This starting point is used by William Waldorf in constructing estimates of labor quality.

[3]See Laurits Christensen, Dale Jorgenson, and Lawrence Lau.

[4]This translog growth rate is derived by Spencer Star and Robert Hall, with the approximation error for arithmetic share weighting of discrete data. The relation between the procedure and translog functions in homogeneous form is derived by W. Erwin Diewert (1976, 1980).

and the growth rate for quality change in discrete time is

$$q_t \equiv d_t - h_t \tag{12}$$

where d_t is defined by (10) and h_t by (11). If hours of relatively high wage labor increase more rapidly than total hours, q_t is positive.

In the labor market, a series of employment characteristics is observable, determining the marginal product of each worker. Suppose there are n factors, and the inputs classified mutually exclusively and exhaustively into the levels of these factors. The growth of labor input is, deleting the time subscript,

$$d = \sum_{i(1)=1}^{I(1)} \cdots \sum_{i(n)=1}^{I(n)} v_{i(1),\ldots,i(n)} \Delta ln\, h_{i(1),\ldots,i(n)} \tag{13}$$

where the factors $j=1,\ldots,n$ are classified into levels $I(1),\ldots,I(n)$. Labor input growth is the weighted sum over cell configurations $i(1),\ldots,i(n)$. The aggregate quality change is decomposable into characteristic sources by the use of partial indices of labor input.[5]

A growth rate of a partial labor input index involves share weighting over a proper subset of the factors $j=1,\ldots,n$. A partial translog growth rate uses the v_i of the factors included in the subset, constructed by arithmetic mean share weighting. Labor is regarded as being homogeneous (i.e., wage rates are identical) over the factors which are *not* included in the subset. If the growth of total hours is subtracted, the contribution to quality change of the factor subset is obtained, assuming (possibly erroneously) that labor is homogeneous over the factors not included in the subset.

The growth rate of the partial index including a single factor is

$$d_1 = \sum_{i(1)=1}^{I(1)} v_{i(1)} \times \Delta ln \left(\sum_{i(2)=1}^{I(2)} \cdots \sum_{i(n)=1}^{I(n)} h_{i(2),\ldots,i(n)} \right) \tag{14}$$

where d_1 is the growth rate and the factors are renumbered such that the considered factor is the first. Unweighted summation of total hours occurs over the last $(n-1)$ factors, and the share weight $v_{i(1)}$ is constructed by summation within the level $i(1)$ over the levels of the remaining factors. The growth rate is of a partial information form in that share weighting occurs only over the levels of the first factor.

The growth rate of a partial index of the first j factors is

$$d_{1,\ldots,j} = \sum_{i(1)=1}^{I(1)} \cdots \sum_{i(j)=1}^{I(j)} v_{i(1),\ldots,i(j)} \times \Delta ln \left(\sum_{i(j+1)=1}^{I(j+1)} \cdots \sum_{i(n)=1}^{I(n)} h_{i(j+1),\ldots,i(n)} \right) \tag{15}$$

where the shares $v_{i(1),\ldots,i(j)}$ are constructed by summation over the remaining $(n-j)$ factors within the cell configuration $i(1),\ldots,i(j)$. Each of the partial index growth rates $d_{1,\ldots,j}$ can be justified in the context of the translog aggregation exactly as in (8)–(10), except for the assumption that the wage restriction

$$w_{i(1),\ldots,i(j),i(j+1),\ldots,i(n)} = w_{i(1),\ldots,i(j),k(j+1),\ldots,k(n)}$$

for any $i(j+1),\ldots,i(n)$ and $k(j+1),\ldots,k(n)$ is required.

By this procedure growth rates and partial indices of quality change for all proper subsets of the labor characteristics are derived. The partial indices provide a structure for the source decomposition of quality change. The main effect of factor i or efficiency adjustment to total hours is the difference between the growth of its single factor index d_i and total hours. This yields

$$q_i = d_i - h \tag{16}$$

where q_i is the main effect of factor i. If q_i is positive, then this main effect augments total hours. Labor input constructed as the sum of hours worked is biased downward.[6]

[5]This partial index is developed and applied to labor input measurement by William Barger. The structure is applied to the explanation of labor input growth in aggregate *U.S.* manufacturing, 1948–66.

[6]The aggregation bias in measuring labor input by total hours in noted in Jorgenson and Zvi Griliches. If $q_i > 0$, labor input is understated, and the difference between output growth and aggregate input growth, total factor productivity, is biased upward.

TABLE 1—QUALITY CHANGE DATA, TWO-FACTOR EXAMPLE

	A. One Factor, Education			
	College (C)		High School (H)	
	Total hours (b.)	Wage/hour ($)	Total hours (b.)	Wage/hour ($)
t	10	5	20	3
t+1	12	7	20	3
	$v_C=(.45+.58)/2=.515$		$\Delta \ln h_C=\ln 12-\ln 10=.1823$	
	$v_H=(.55+.42)/2=.485$		$\Delta \ln h_H=\ln 20-\ln 20=0$	
	$d_1=.515\times.1823=.0939$		$h=\ln 32-\ln 30=.0645$	
	B. Two Factors, Education and Age			
	College, Under 30 (C, U)		College, Over 30 (C, A)	
	Total hours (b.)	Wage/hour ($)	Total hours (b.)	Wage/hour ($)
t	2	5	8	5
t+1	3	7	9	7
	High School, Under 30 (H, U)		High School, Over 30 (H, A)	
t	5	2	15	3.33
t+1	4	2	16	3.25
	$v_{C,U}=.12$		$\Delta \ln h_{C,U}=.4055$	
	$v_{C,A}=.40$		$\Delta \ln h_{C,A}=.1178$	
	$v_{H,U}=.07$		$\Delta \ln h_{H,U}=-.2231$	
	$v_{H,A}=.41$		$\Delta \ln h_{H,A}=.0645$	
	$v_U=.19$		$\Delta \ln h_U=0$	
	$v_A=.81$		$\Delta \ln h_H=\ln 25-\ln 23=.0834$	
	$d_2=.81\times.0834=.0675$			
	$d_{12}=(.12\times.4055)+(.40\times.1178)+(.07\times-.2231)+(.41\times.0645)=.1067$			

If there are two factors i and k as a proper subset from n factors, a first-order interactive effect is derived from the partial index growth rate d_{ik} for the two factors and the single factor indices d_i and d_k. The first-order interactive effect between i and k is

$$(17)\quad q_{ik}=(d_{ik}-h)-(d_i-h)-(d_k-h)$$
$$=d_{ik}-h-q_i-q_k$$

as the total joint effect of i and k or $(d_{ik}-h)$ less the main effect of each. If there are only two factors, quality change is

$$(18)\quad q=d_{ik}-h$$
$$=q_i+q_k+q_{ik}$$

the sum of the main effects and the interaction. From the $d_{1,\ldots,j}$, interactive effects up to $(j-1)$th order are obtainable. In a four-factor classification, effects up to third order are derived, and quality change is

$$(19)\quad q=d_{1,\ldots,4}-h$$
$$=\sum_{i=1}^{4} q_i+\sum_{i=1}^{4}\sum_{j=i+1}^{4} q_{ij}$$
$$+\sum_{i=1}^{4}\sum_{j=i+1}^{4}\sum_{k=j+1}^{4} q_{ijk}+q_{1234}$$

where q_{1234} is the third-order interaction.

A numerical example detailing the procedure and errors from incorrect specification is given in Table 1. Panel A presents a classification by education when age, another determinant of labor input, is omitted. Labor input grows at 9.39 percent over the period, and $q_1=d_1-h=.0294$ is the education effect. When the correct and complete data of panel B are used labor input growth is d_{12} or 10.67 percent, so a downward bias of 13.6 percent arises from incorrectly excluding age.

The analogous age effect is $q_2=d_2-h=$ .0030, so the shifting age composition increases labor input by 0.3 percent. The

education-age interaction is $q_{12}=d_{12}-d_1-d_2-h=.0098$, indicating that this effect adds to labor input. The sources of growth in labor input are total hours worked, 6.45 percent; changing educational composition, 2.94 percent; age composition, 0.03 percent; and education-age interaction, 0.98 percent. With a linear logarithmic aggregate, the interaction would be restricted to zero, and labor input growth biased downward, relative to d_{12}, by 9 percent.[7]

More generally, interactions in an n-factor classification are constructed analogously as marginal contributions given total hours, all main effects, and all interactions up to order $n-1$. Quality change is a growth accounting identity, being the sum of main effects for each factor and interactive effects involving subsets of factors.

II. Data Construction

The structure is applied to a complete crossed classification of total hours worked and average wages per hour in the *U.S.* private domestic economy 1947–74. There are 1,600 types of labor, indicated by factor and levels in Table 2. There are five educational groups, eight age groups, ten occupations (one digit), and two classes of worker for each sex. The data are described briefly here, with more detail contained in Frank Gollop and Jorgenson.

The data are controlled to match series on total jobs filled and hours worked published by the Bureau of Economic Analysis (*BEA*) of the U.S. Department of Commerce in its *Survey of Current Business*. These data are based on payroll records reported to state unemployment insurance bureaus, but contain no information on the relevant employment characteristics.

These characteristics of class, sex, age, education, and occupation, together with industry, are obtained from household surveys of the labor force, published in the decennial Census and the *Current Population Survey* (*CPS*). The industry classification is almost identical to that for the *BEA*, and contains fifty-one sectors, with 1,600 types of labor in each, or 81,600 groups. Published cross tabulations are available for subsets of the factors, but not the entire classification. The objective is to combine the available partial classifications to construct the 1,600 types of labor for each industry and year.

The procedure is summarized in Table 3. All partial classifications on persons employed for each year are obtained, for example, two-way sex-age or sex-occupation tables. These are multiplied together using a multivariate version of the biproportional algorithm of Michael Bacharach, and controlled to *BEA* jobs filled. This yields series on jobs filled for all 1,600 labor groups.

Hours worked per job are published as frequency distributions by the Census and *CPS*. Mean hours worked per job are estimated for all available partial classifications. The jobs-filled series is summed to be consistent with these mean hours. If mean hours worked by sex and occupation are

TABLE 2—CLASSIFICATION OF LABOR INPUT

Education
Elementary (0–8 years of elementary school)
Some high school (1–3 years of high school)
High school graduate
Some college (1–3 years of college)
College graduate (4 years or more of college)
Sex
Male, Female
Age
14–15, 16–17, 18–24, 25–34, 35–44, 45–54, 55–64, 65 and above
Class of Worker
Private wage and salary employee
Self-employed or unpaid family worker
Occupation
Professional, technical and kindred
Managerial
Farm and farm managerial
Clerical
Sales
Service (including private household workers)
Operative
Crafts
Farm laborer
Laborer

[7]A version of this noninteractive index is applied to labor input measurement by Edward Denison, for the *U.S.* business sector 1929–69. The form used contains the factors class of worker, sex, age, and education. A sex-age interaction is permitted, but others are restricted to zero.

TABLE 3—LABOR INPUT DATA CONSTRUCTION BY INDUSTRY

			Source
	(1)	**Employment**	
		Persons employed (partial)	Census, *CPS*
Control to		Jobs filled	*BEA*
=		Jobs filled (1,600 types) (1)	
	(2)	**Hours worked**	
		Mean hours worked/job (partial)	Census, *CPS*
×		Jobs filled (see (1))	
=		Total hours worked (partial)	
Control to		Total hours worked	*BEA*
=		Total hours worked (1,600 types)	
Divide by		Jobs filled (1,600 types) (1)	
=		Mean Jobs/Worked/Job (1,600 types)	
	(3)	**Weeks paid: Jobs per person**	
		Mean weeks paid/worker (partial)	Census, *CPS*
Divide by		Fifty-Two Weeks	
=		Mean Jobs/worker (partial)	
×		Jobs filled (see (1))	
Multiply Together			
=		Total Jobs/worker (1,600 cells) (3)	
Multiply by		Fifty-Two Weeks	
=		Mean weeks paid/worker (1,600 cells) (3)	
	(4)	**Hourly user cost of labor (employees)**	
		Mean earnings/worker-year (partial)	Census, *CPS*
Add		Social Security and unemployment	
=		Mean adjusted earnings/worker-year (partial)	
Divide by		Total Jobs/Worker (1,600 cells) (3)	
=		Mean adjusted earnings/job (partial)	
×		Jobs filled (see (1))	
=		Total Adjusted Earnings (partial)	
Control to		Wages Plus Supplements	*BEA*
=		Total Earnings (1,600 cells)	
Divide by		Hours Worked/Job (see (2))	
X		Jobs Filled (see (1))	
Multiply by		Fifty-Two Weeks	
=		Mean User Cost Per Hour	

available, total hours worked are obtained by multiplying jobs filled summed over class, education and age, and mean hours. These are controlled to *BEA* series on total hours worked.

A job is any activity providing fifty-two weeks of paid employment during a year. What are reported in household surveys as weeks worked include sick leave and vacation, so the actual series is weeks paid. The procedures for constructing series on weeks paid are analogous to those for hours worked per job except that there is no *BEA* control total on weeks paid by industry. Jobs per worker are weeks paid divided by fifty-two given the definition of a job.

It is assumed that marginal products of employees are equated with the user cost of labor. The user cost is the wage rate plus any hourly related wage supplements. Means are estimated for earnings frequency distributions for employees, with the algorithm containing a tax calculator to add employer contributions to Social Security and payroll unemployment insurance. The biproportional procedure is used to construct total earnings for all 800 types of employees, inclusive of employer supplements. This total is divided by the product of hours worked per job, jobs filled, and fifty-two to yield the mean user cost of labor.

For the 800 self-employed cells, total labor compensation must be imputed. The after-tax rate of return on capital is assumed equal in the corporate and noncorporate sectors. This return is multiplied by noncorporate capital stock, yielding an estimate of noncorporate capital income. This is subtracted, by industry, from noncorporate income to yield noncorporate labor income. Since hours worked per job and jobs filled are estimated from Table 3, imputed hourly wage rates for all 800 cells can be constructed.

The relevant wages $w_i, i=1,\ldots,1{,}600$ are the user costs of labor for employees and the self-employed, with compensation, hours, and employment summed over industry. Total hours h_i are the product of jobs filled, mean hours worked per job and fifty-two. Labor compensation shares are

$$s_i = w_i h_i \Big/ \sum_{i=1}^{1600} w_i h_i$$

calculated annually for 1947–74. The v_i are constructed as two-year moving averages of the s_i.

III. Empirical Results

The results of the quality change decomposition are presented and the magnitudes of factor effects discussed and compared with alternative estimates. The sum of the interactions is examined to determine whether labor quality can be represented by a main effects or linear logarithmic index. Marginal quality contributions are computed for specific factors, given the four remaining factors, and these are compared with the main effects.

The quality change decomposition is indicated in Table 4 for all effects and interactions. Quality change is the sum of five main effects, and interactions comprising ten of first order, ten of second order, five of third order, and one of fourth order, for a total of thirty-one. For the period 1947–51 quality change or growth in labor input per hour worked averages 1.12 percent per annum. Quality change exhibits fluctuations over the period, declining to −.23 percent during 1963–67, largely because of substantial negative main effects attributable to age and sex. Since negative effects arise if low-wage employment increases more rapidly than high-wage employment, this is attributable to increasing shares of total hours worked for young workers and women.

The linear logarithmic quality growth rate is the sum of main effects only, and averages 0.90 percent over the period 1947–74, as opposed to 0.60 percent when interactions are taken into account. This creates an upward bias of 50 percent if the linear logarithmic specification is used, overstating the rate of growth of labor input. Comparing the main effects, the overall contribution of sex composition is to reduce growth in labor input per hour by 0.15 percent per annum. Spencer Star has obtained an estimate of −.01 percent for an analogous main effect of sex for 1950–60.[8] In the period 1951–59, the quality change sex effect is −.26 percent. Barger obtains a main effect for sex for *U.S.* manufacturing, 1948–66, of −.24 percent, similar to the derived result. Both measures are based on similar arithmetic share weighting.

A watershed develops in the main effect of age within the period, as after 1955–59 age becomes a retarding factor. The increase in the share of total hours worked by younger workers decreases the main effect, but for the period as a whole the effect of age is close to zero. The main effect of class is positive at .15 percent, accounting for the decline in the relative share of hours worked by the self-employed. Denison estimates that for the *U.S.* business sector, 1950–62, the movement of all resources from the noncorporate to corporate sectors increases output per capita by .29 percent, not inconsistent with the obtained results.[9]

[8]Labor factors are added in a specified order to a total hours index. Since sex is the first included, the measure is the sex effect. The estimate exceeds that in Table 3, but labor is distinguished by employment as opposed to total hours worked.

[9]Denison constructs imputed noncorporate income using a scale factor based on employee wages and corporate after-tax returns applied to labor and capital, respectively.

TABLE 4—QUALITY CHANGE DECOMPOSITION, *U.S.* PRIVATE DOMESTIC ECONOMY 1947–74
(Average Percent Per Year)

	1947–74	1947–51	1951–55	1955–59	1959–63	1963–67	1967–71	1971–74
Main Effects								
Sex (*S*)	−.15	.04	−.19	−.34	−.05	−.24	−.22	−.06
Class (*C*)	.14	.28	.13	.26	.14	.04	.05	.09
Age (*A*)	−.01	.34	.27	.03	−.07	−.22	−.20	−.29
Education (*E*)	.61	.27	.50	.58	.72	.85	.81	.67
Occupation (*J*)	.31	.63	.32	.42	.37	.14	.40	−.11
	.90	1.56	1.03	.95	1.11	.57	.84	.30
Interactive Effects								
First Order								
SC	.05	.05	.00	.01	.14	.17	.00	.02
SA	.02	−.07	−.02	−.03	.13	.12	.02	−.01
SE	.07	.03	.06	.09	.13	.13	.03	.02
SJ	.09	−.01	.12	.17	.17	.15	.07	−.03
CA	.04	.07	.06	.02	.12	.04	−.01	−.03
CE	−.05	−.03	−.04	−.05	.06	−.20	−.04	−.02
CJ	−.10	−.10	−.06	−.12	.07	−.51	.03	−.05
AE	.02	.01	.05	.00	.12	.03	−.01	−.07
AJ	.00	−.10	−.04	−.04	.09	.00	.04	.06
EJ	−.27	−.22	−.30	−.41	−.18	−.36	−.35	−.05
	−.13	−.37	−.17	−.36	.85	−.43	−.22	−.16
Second Order								
SCA	−.03	−.01	−.01	−.01	−.09	−.07	−.00	−.01
SCE	−.03	.00	.01	.00	−.10	−.11	−.00	−.01
SCJ	−.06	−.06	−.00	−.01	−.15	−.20	.01	−.01
SAE	−.02	−.01	−.02	−.02	−.09	.00	.02	.01
SAJ	−.02	.02	−.01	−.01	−.09	−.07	.00	−.01
SEJ	−.07	−.03	−.03	−.06	−.17	−.12	−.05	−.04
CAE	.00	.01	.03	.03	−.06	−.02	.02	.02
CAJ	−.03	−.04	−.03	−.02	−.11	.00	−.01	.01
CEJ	.01	.02	.04	.05	−.08	.04	−.01	−.02
AEJ	−.01	.02	.03	.05	−.12	−.05	−.01	.02
	−.27	−.08	−.01	.00	−1.06	−.60	−.03	−.04
Third Order								
SCAE	.02	.01	.00	.00	.11	.03	.00	.00
SCAJ	.02	.01	.01	.01	.10	.04	.00	.01
SCEJ	.03	.00	−.01	.00	.10	.13	.00	.00
SAEJ	.03	.01	.02	.02	.10	.03	−.01	.01
CAEJ	.01	−.01	−.02	−.02	−.02	.07	.03	.00
	.11	.02	.00	.01	.48	.26	−.01	.02
Fourth Order								
SCAEJ	−.02	−.01	.00	.00	−.11	−.03	.00	.00
Quality Change	.60	1.12	.85	.60	1.27	−.23	.58	.12
Total Hours	.86	.59	.29	.29	−.03	2.54	.26	2.55
Labor Input	1.46	1.71	1.14	.89	1.24	2.31	.84	2.67

The education effect increases monotonically to .85 percent in 1963–67, but with the decline in relative wages of the college educated outweighing their increasing share of total hours, the effect is .67 percent for 1971–74. A single factor education index in Christensen and Jorgenson yields .71 percent for the *U.S.* private domestic economy, 1947–69. Denison derives an estimate of .65 percent for the *U.S.* business sector, 1947–69.[10] These estimates are similar to the result obtained. The latter measure includes a correction for the quality of schooling by the use of length of school year as a classifying variable.

[10]The educational index weights distributions of employment by earnings relatives, assuming perfect substitutability between educational levels.

Waldorf obtains an estimate of .32 percent for labor quality change in *U.S.* manufacturing by occupation, 1952–67, using arithmetic share weighting. The occupational effect becomes negative at −.11 percent for 1971–74, indicating that the growing job opportunities in this period occur in those with relatively low wages.

Among the first-order interactions, that for sex and age is negligible, indicating a proportionate entry of women in all age groups. The remaining age interactions with class, education, and occupation are all close to zero. Consequently, the relative increase in the proportion of total hours worked by younger workers is not concentrated in any given skill class. In no case is any age interaction of substantial magnitude.

The interactions of sex with education and occupation are, respectively, over 1947–74, .07 and .09 percent, implying that the proportion of women has increased more rapidly in skilled labor categories. These two first-order interactions alone eliminate the entire main effect of −.15 percent per year, and amount to over one-quarter of the 1947–74 quality change. Moreover, the interaction with class is positive at .05 percent per year, implying that women have shifted more rapidly than men from self-employment, including unpaid family workers, to employees. Quality change constructed by main effects summation overstates the reduction in labor input through increases in female employment.

The class interactions with education and occupation are both negative, indicating that employees are concentrated in high-wage, education, and occupation groups. The education and occupation interaction is negative, at −.27 percent over the period. High-wage groups, one–three years of college and college graduates in educational attainment, are also in skilled occupations. The interaction reduces the effect of education and occupation from 0.92 percent per annum to 0.65 percent.

The interactions of higher order are of small magnitude but in sum are consistently negative. The largest effect is the education-sex-occupation interaction, given that high-wage education-sex groups have high proportions of total hours worked by persons in relatively high-wage occupations.

The last three entries in the rows of Table 4 correspond to q_t, h_t, and d_t, the growth rates of labor input per hour worked, total hours worked, and labor input. There is a role reversal between quality change and total hours as sources of growth during the period. The average quality change in the demobilization period 1947–51 amounts to 1.12 percent with a positive sex effect associated with increasing male employment share. This rate of growth declined to .12 percent by 1971–74, mainly attributable to rapid growth in low-wage occupations and shifts towards a younger work force.[11] The 1959–63 period marks the turning point. During 1947–63 the principal source of labor input growth is quality change, while for 1963–74 total hours become dominant. Hours worked exhibit a relatively slow increase up to 1963 and a rapid increase thereafter. Whereas for 1947–74 quality change accounts for 31 percent of the growth in labor input, the proportion is less than 4 percent after 1971.

The sum of the main effects elements yields the estimate of quality change consistent with linear logarithmic aggregation. The use of such a form overstates the quality measure, with the exception of the period 1963–67. When quality change is positive, the main effects growth rate exceeds total quality change, implying that a linear logarithmic aggregation yields upward biases. The major interactive effects are those of second order, and the total of these has sign opposite to that of the main effects sum. This is a consequence of positive correlation between factor effects.

The marginal effect is the change in the labor input growth rate formed from $(n-1)$ factors when the nth is added. The interactive effect is the sum of interactions of all orders containing this factor, and is the difference between the marginal and main

[11]In Barger, an aggregate for *U.S.* manufacturing by sex, age, and education is reduced from 1948–60 to 1960–66. Waldorf, using an occupational classification for *U.S.* manufacturing, obtains a decrease from 1952–60 to 1960–67.

TABLE 5—FACTOR CONTRIBUTIONS TO QUALITY CHANGE
(Percent Per Year)

	1947–51	1951–55	1955–59	1959–63	1963–67	1967–71	1971–74	1947–74
Sex								
Marginal	−.07	−.09	−.18	.12	−.05	−.14	−.08	−.07
Interactive	−.11	.10	.16	.17	.19	.08	−.02	.08
Class								
Marginal	.17	.09	.13	.19	−.61	.04	.00	.01
Interactive	−.11	−.04	−.13	.05	−.65	−.01	−.09	−.13
Age								
Marginal	.23	.29	.01	.08	−.15	−.16	−.21	.02
Interactive	−.11	.02	−.02	.15	.07	.04	.08	.03
Education								
Marginal	.05	.30	.25	.49	.38	.39	.41	.34
Interactive	−.22	−.20	−.33	−.23	−.47	−.42	−.26	−.27
Occupation								
Marginal	.20	.28	.08	.05	−.79	.10	−.15	−.09
Interactive	−.43	−.04	−.34	−.32	−.93	−.30	−.04	−.41

effects. The results of this procedure are presented in Table 5 where the marginal effect is the first entry, and the sum of the interactive effects the second entry. The main effect shown in Table 4 plus the interactive effect is the marginal effect. The marginal effect for sex exceeds the main effect, given the above average increase in women in skilled labor types indicated by the positive interaction effect. Main effects or linear logarithmic analysis of sex composition as a source of quality change overstates the decline in quality associated with the entry of women. The increase in the proportion of total hours worked by women increases the skill content of the average hour, given the positive sex interactions with education and occupation. The effect of the adjustment for interactive effects is to reduce the negative contribution of sex from −.15 to −.07 percent, or about one-half, over the entire period.

The marginal effect of class for 1947–74 is close to zero, implying that the relative decline of self-employment is captured by the levels of the remaining four characteristics. The age marginal effect is also close to zero for the entire period, but declining subsequent to the watershed period 1959–63. The interactive effect of age is negligible throughout, suggesting that linear logarithmic aggregation with a main effect only is appropriate for this factor.

The marginal effect of education is .34 percent, the additional contribution of this factor to an index already containing sex, class, age, and occupation. This is about one-half the main effect of education, implying that this portion is measurable by a linear combination of the remaining characteristics. The marginal effect of occupation is substantially less than the main effect. The inclusion of sex, class, education, and age interaction reduces the contribution of occupation from .32 to −.09 percent over the period.

The measured effects from altering the sequence of entering factors differ considerably. This is because of the magnitude of the interaction terms. If education is added to the index after occupation, the negative effect of the first-order interaction reduces the contribution of education to labor input.

IV. Summary and Conclusions

Estimates of sources of quality change in labor input have been derived using the linkage results between specification of the underlying functional form, index numbers corresponding to the form, and separability of labor from other factors of production. The separability assumption permits construction of measures of labor productivity without recourse to interaction between labor and nonlabor inputs. Such separability

should be tested in expansion of consideration to nonlabor inputs. The functional form used, the translog, has the property of approximate consistency in aggregation. This implies that little error arises from a two-stage construction of value-added by forming subaggregates of labor and nonlabor inputs, and subsequently aggregating the two. This vitiates in large part any error from incorrect aggregation if separability does not obtain. Further work is required on the structure of the effects themselves. While no statistical testing is possible, the number of effects can be truncated at the first-order level if the aggregate of effects is translog. This yields a complete theory of labor productivity and hedonic labor indices connected with appropriate functional forms.

It has been demonstrated that the effect of increased female hours relative to males subsequent to World War II has been overstated in its reduction of labor quality. The results obtained indicate a negligible negative contribution before adjustments of relative wages by sex for discrimination or imperfect information.

The underlying contribution of education is reduced by one-half once adjustments are made for skill composition. Moreover, the main effect of education, although positive, declines by one-fifth from the early 1960's to the early 1970's. This reduction of .18 percent in the main effect of education would alone reduce output growth by .13 percent per year for a labor share of two-thirds. The evidence is that the large increases in employment shares of the relatively educated are failing to maintain the rise in productivity growth of the postwar era. Further examination of the return from investment in education is required.

Difficulties are posed in constructing labor input data consistent with the national accounts, and these problems are by no means resolved. The *CPS* survey week may be unrepresentative of the year. Wage data cannot accurately be attributed to each type of labor, and the biproportional procedure imparts an inflexibility in the allocation of labor which requires empirical testing.

The results depend on which classification is used, and the 1,600 types of labor are not exhaustive. Also, further work is required on whether user costs and marginal products are equated in the labor market, and whether observed wages contain market distortions.

REFERENCES

M. Bacharach, "Estimating Non-Negative Matrices from Marginal Data," *Int. Econ. Rev.*, Sept. 1965, *6*, 294–310.

W. J. Barger, "The Measurement of Labor Input: U.S. Manufacturing Industries 1948–1966," unpublished doctoral dissertation, Harvard Univ. 1971.

L. R. Christensen and D. W. Jorgenson, "Measuring Economic Performance in the Private Sector," in Milton Moss, ed., *The Measurement of Social and Economic Performance*, New York 1973.

———, ———, **and L. J. Lau,** "Transcendental Logarithmic Production Frontiers," *Rev. Econ. Statist.*, Feb. 1973, *55*, 28–45.

Edward. F. Denison, *Accounting for United States Economic Growth, 1929–1969*, Washington 1974.

W. E. Diewert, "Exact and Superlative Index Numbers," *J. Econometrics*, Jan. 1976, *4*, 115–45.

———, "Aggregation Problems in the Measurement of Capital," in Dan Usher, ed., *The Measurement of Capital*, New York forthcoming 1980.

F. M. Gollop and D. W. Jorgenson, "U.S. Productivity Growth by Industry, 1947–1973," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement*, New York forthcoming 1980.

D. W. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Rev. Econ. Stud.*, July 1967, *34*, 249–83.

R. R. Nelson, "Recent Exercises in Growth Accounting: New Understanding or Dead End?," *Amer. Econ. Rev.*, June 1973, *63*, 462–68.

S. Star, "Accounting for the Growth of Out-

put," *Amer. Econ. Rev.*, Mar. 1974, *64*, 123–35.

______ and R. E. Hall, "An Approximate Divisia Index of Total Factor Productivity," *Econometrica*, Mar. 1976, *44*, 257–64.

W. Waldorf, "Quality of Labor in Manufacturing," *Rev. Econ. Statist.*, Aug. 1973, *55*, 284–91.

U.S. Bureau of Labor Statistics, *Current Population Survey* (*CPS*), Washington, various years.

U.S. Office of Business Economics, (*BEA*) *Surv. Curr. Bus.*, Washington various years.

Efficient Foreign Exchange Markets and the Monetary Approach to Exchange-Rate Determination

By DOUGLAS W. CAVES AND EDGAR L. FEIGE*

Recent contributions to the balance-of-payments literature (see Robert Barro; Rudiger Dornbusch; Jacob Frenkel; Hans Genberg and Henry K. Kierzkowski; Donald Kemp; Michael Mussa; Michael Parkin) have focused on the determination of the exchange rate viewed as a relative price of two national monies. We shall refer to this approach as the monetary approach. From this perspective, the exchange rate is seen as endogenously determined by stock equilibrium conditions in markets for national monies, and the major theoretical focus is the supply and demand functions for monetary stocks. The demand functions are typically specified to correspond to conventional macro-model money demand equations involving income, the rate of interest, and, perhaps most importantly, expectations concerning future exchange rates. Several of the specifications permit the simultaneous determination of the exchange rate and the expected future exchange rate by imposing restrictions which characterize rational expectation formation. Other models include broader specifications, and additional asset supply and demand equations. Despite their differences, a common theme runs through all these studies: While it is recognized that the exchange rate is not purely a monetary phenomenon (being affected in part by the real determinants of the demand for money), the thrust of the new asset approach to exchange rates is that the exchange rate is primarily determined by current and past innovations in the relative supplies of national monies. It is therefore not surprising that the few empirical efforts at testing this approach to exchange-rate determination have estimated reduced-form equations for the exchange rate as a distributed-lag function of the domestic money supply or the ratio of the domestic money supply to the foreign money supply.[1] These empirical tests implicitly assume exogeneity of the money supply in order to assure consistency of dynamic regression estimates.

At the same time that we are witnessing an elaboration of monetary theories of exchange-rate determination, there exists a parallel literature (see Michael Dooley and Jeffrey Shafer; William Poole; John Pippenger; John Geweke and Feige) which develops a theory of exchange-rate determination based on the hypothesis that foreign exchange markets are efficient markets. These writings are largely derivative from the extensive theoretical developments in the security market literature (see Eugene Fama). The efficient market hypothesis suggests that economic agents act upon information available to them in such a way that the spot exchange rate always reflects all available information that could be potentially useful in developing trading rules that earn excess profits.

To date, tests of the efficient foreign exchange market hypothesis have been limited

*Graduate student and professor of economics, respectively, University of Wisconsin-Madison. We wish to thank Robert McGee for computational assistance and George Borts and John Geweke for many helpful comments on an earlier draft. The research was supported by grants from the Graduate School of the University of Wisconsin and National Science Foundation grant SOC 76-21439.

[1]Frenkel estimates a polynomial distributed lag of the effects of the money supply on the exchange rate for the German hyperinflation whereas Stephen Magee reports the results of some unpublished papers which utilize the ratio of two money stocks. Similarly, John Bilson and Neil Sargent include both money supplies as determinants of the exchange rate.

to "weak form" tests, which simply examine the hypothesis that all information contained in the past history of exchange rates is already reflected in the current rate. The broader theory of efficient markets, however, recognizes that efficiency is contingent upon the specific information set which conditions expectations of future prices. Once information in addition to the history of exchange rates is permitted to enter the set upon which investor expectations may be conditioned, a variety of efficiency notions can be developed. To order these notions in a manner which is meaningful in assessing both the market efficiency propositions and the simple monetary propositions, we introduce the concept of incremental efficiency. The foreign exchange market is said to be incrementally efficient with respect to some type of information if the use of that information cannot improve upon forecasts of the exchange rate which are conditional only on the history of the exchange rate. If the foreign exchange market is incrementally efficient with respect to monetary disturbances, then the spot exchange rates already incorporate the systematic information contained in the past history of money supply disturbances, thus vitiating some of the directly testable implications of the monetary approach to exchange-rate determination.

In spite of the fact that proponents of the monetary approach have incorporated rational expectation formation specifications in their models and have on occasion referred to the efficient market literature, they have failed to recognize that one of the consequences of an efficient foreign exchange market is to eliminate the possibility of directly observing a systematic relationship exchange rates and past supplies of national monies. If the foreign exchange market is efficient, all monetary effects on exchange rates will be contemporaneous. Thus, while monetary theories and efficient market theories are by no means conceptually inconsistent with one another, the existence of efficient markets may severely restrict the nature of the direct tests of the monetary approach to exchange-rate determination.

The primary purpose of this paper is to formulate a simple analytic framework that is capable of characterizing the major hypotheses embodied in both the monetary approach to exchange-rate determination and the efficient market approach. To this end we present a bivariate stochastic model of exchange-rate changes and relative money supply changes, and apply the "causal detection" analysis proposed by C. W. J. Granger and Christopher Sims (1972) in order to empirically test the salient assumptions and predictions of both theories. The bivariate stochastic representation, while abstracting from many of the more complicated specifications of exchange-rate determination, is nonetheless sufficiently rich to allow investigation of the most salient propositions of the simple monetary and efficient market approaches to exchange-rate determination.

The market efficiency propositions we investigate are:

PROPOSITION E1: *That the market for foreign exchange can be characterized as being weak-form efficient.*

PROPOSITION E2: *That the market for foreign exchange is incrementally efficient with respect to the supplies of national monies.*

PROPOSITION E3: *That both Propositions* E1 *and* E2 *are jointly acceptable, implying that the market displays semistrong-form efficiency.*

We also investigate the following propositions relating to the simple monetary approach:

PROPOSITION M1: *That changes in the stocks of national monies are exogenous with respect to variations in exchange rates.*

PROPOSITION M2: *That changes in exchange rates are not caused by past variations in national money supplies; that is, that the most basic assertion of the simple monetary approach is not supported by the data.*

PROPOSITION M3: *That the sole relationship between money supplies and exchange rates is contemporaneous; that is, that the simple monetary approach may be valid but that the existence of an efficient market forces the entire impact of the money supply to be instantaneous.*

Proposition M1 addresses the issue of the existence of a government reaction function whereby the central banks intervene in foreign exchange markets in order to smooth out fluctuations in exchange-rate movements.[2] If such intervention is responsive to exchange-rate movements, then the typical assumption of monetary exogeneity is invalid. Thus, rejection of Proposition M1 suggests that efforts to estimate dynamic regressions using least squares methods as reported, for instance, by Frenkel and Magee may yield inconsistent parameter estimates. Confirmation of Proposition M2 is evidence against what we regard to be the major substantive hypothesis of the monetary approach to exchange-rate determination, namely, that changes in national money supplies are the primary determinant of exchange-rate movements under regimes of flexible exchange rates and regimes of controlled floating exchange rates. Evidence consistent with Proposition M3 supports the *joint* hypothesis of efficient foreign exchange markets and monetary determination of exchange rates.

Section I briefly summarizes a simplified model of the monetary approach to exchange-rate determination under a regime of flexible exchange rates, and presents a generalized bivariate stochastic model of the relationship between the exchange rate and the money supply. Section II develops and tests the restrictions on the bivariate stochastic process which corresponds to Propositions E1, E2, and E3. In Section III the restrictions corresponding to Propositions M1, M2, and M3 are developed and tested. The final section presents a summary of the findings and suggestions for further research.

I. Causality Relationships in a Bivariate Model of Exchange-Rate Determination

To date, there exists no single model which could be represented as "the" monetary model of exchange-rate determination, and it is well beyond the purview of this paper to attempt to specify a particular structural model which would enjoy a consensus even among advocates of the monetary approach. Instead, we will content ourselves with a brief presentation of some of the salient features which recent exchange-rate models incorporate, and focus attention on the forecasting equations for exchange-rate changes and relative supplies of national monies. The emphasis on the forecasting equation for exchange-rate changes also permits the integration of efficient market concepts which have not as yet been incorporated in most monetary approaches.

The monetary approach views the exchange rate as a relative price of two national monies and therefore focuses attention on the specification of the stock demand and supply functions for the national monies. The money supply of a particular country is typically specified as a function of foreign exchange reserves and domestic credit. The arguments of the money demand function typically include interest rates, income, a world price level, the exchange rate, and the expected rate of return on world money relative to that on domestic money.[3] This last (usually expressed as the expected rate of change in the exchange rate) takes on a special importance for the dynamic evolution of the exchange rate, and recent models focus on the simultaneous determination of both the exchange rate and the expected future exchange rate by imposing the restriction that expectations are formed rationally.

[2]Pippenger and Llad Phillips extensively analyze a model which allows for governmental intervention. This reaction function is typically suppressed in monetarist models which generally assume that the money supply is exogenous with respect to the exchange rate.

[3]See, for example, Barro; Dornbush; Frenkel; Genberg and Kierzkowski; Kemp; Mussa; Parkin.

Under fixed-exchange-rate regimes, the monetary authority will be forced to intervene in the exchange market in order to offset appreciations or depreciations resulting from disturbances in the arguments of the money demand function. Under fully flexible-exchange-rate regimes, the money supply is specified to be exogenous with respect to exchange-rate variations, while controlled floating regimes must allow for the specification of an endogenous money supply component to reflect the active interventions in foreign exchange markets required to stabilize exchange rates within desired limits.[4] Given these considerations, it is important that our analytic framework be rich enough to encompass monetary exogeneity and monetary endogeneity (government reaction functions) as special cases.

The major thrust of the monetary approach to exchange-rate determination involves the assertion that exchange-rate fluctuations are largely explicable in terms of variation in the relative supplies of national monies. The monetary approach thus suggests that the money supply could be used to forecast exchange rates; that is, that there exists an observable causal relationship between exchange rates and money supply changes. In contrast, the existence of an incrementally efficient market suggests that past monetary disturbances (to the extent that they portend future monetary shocks) are already reflected in spot rates. Thus, it is important that our analytic framework be capable of including the money stock (either domestic or both domestic and foreign) among the determinants of the exchange rate.

An analytic framework which is ideally suited for representing all of the special cases of interest in the exchange-rate/money supply relationship is the bivariate stochastic process:

$$(1)\qquad \begin{bmatrix} M_t \\ e_t \end{bmatrix} = \begin{bmatrix} \alpha(B) & \beta(B) \\ \gamma(B) & \delta(B) \end{bmatrix} \begin{bmatrix} \varepsilon_{Mt} \\ \varepsilon_{et} \end{bmatrix}$$

[4]Mussa discusses the money supply function under alternative regimes. His structural specification is limited to an exogenously determined money supply. Pippenger and Phillips allow for a richer money supply specification. See fn. 2.

where M_t is the growth rate of the money stock in period t and e_t is the growth rate of the exchange rate in period t. The matrix elements $\alpha(B)$, $\beta(B)$, $\gamma(B)$, and $\delta(B)$ are polynomials in the lag operator B;

$$\alpha(B) = \sum_{i=0}^{\infty} \alpha_i B^i$$

and $$B^n(X_t) = X_{t-n}$$

ε_{Mt} and ε_{et} are serially uncorrelated random variables with zero means and contemporaneous covariance matrix Σ. We impose the identifying restrictions that $\alpha_0 = \delta_0 = 1$, and $\gamma_0 = \beta_0 = 0$. The attraction of the bivariate stochastic model lies in its generality. Equation (1) implies no substantive restrictions on the exchange-rate/money supply relationship; yet empirically testable restrictions on the elements $\alpha(B)$, $\beta(B)$, $\gamma(B)$, and $\delta(B)$ yield formulations which correspond exactly to the money stock exchange-rate relationships envisioned in alternative economic theories.[5] The bivariate stochastic process is thus a powerful tool for choosing among competing theories of exchange-rate determination.

Additional insight into the bivariate stochastic system depicted in (1) is obtained from the unrestricted forecasting equations for e_t and M_t. To derive forecasting equations we invert the parametric matrix in (1), to obtain

$$(2)\qquad \frac{\begin{bmatrix} \delta(B) & -\beta(B) \\ -\gamma(B) & \alpha(B) \end{bmatrix}}{[\alpha(B)\delta(B) - \gamma(B)\beta(B)]} \begin{bmatrix} M_t \\ e_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{Mt} \\ \varepsilon_{et} \end{bmatrix}$$

Defining $H(B) = [\alpha(B)\delta(B) - \gamma(B)\beta(B)]$, we can write the forecasting equations of the system as

$$(3a)\qquad M_t = d(B)M_{t-1} + b(B)e_{t-1} + \varepsilon_{Mt}$$

$$(3b)\qquad e_t = a(B)e_{t-1} + c(B)M_{t-1} + \varepsilon_{et}$$

[5]See, for example, Mussa and Barro for explicit versions of these restrictions, and Frenkel for an implicit monetary exogeneity assumption.

where

$$(4a)\quad Bc(B) = \sum_{i=0}^{\infty} c_i B^{i+1} = H^{-1}(B)\gamma(B)$$

$$(4b)\quad [1 - Ba(B)] = \left(1 - \sum_{i=0}^{\infty} a_i B^{i+1}\right) = H^{-1}(B)\alpha(B)$$

$$(4c)\quad Bb(B) = \sum_{i=0}^{\infty} b_i B^{i+1} = H^{-1}(B)\beta(B)$$

$$(4d)\quad [1 - Bd(B)] = \left(1 - \sum_{i=0}^{\infty} d_i B^{i+1}\right) = H^{-1}(B)\delta(B)$$

The unrestricted forecasting equations in (3) display the money supply process as a function of the past history of the money supply, past exchange-rate changes, and a contemporaneous random shock. The change in the exchange rate is shown to depend on its own past history, past money supply changes, and a contemporaneous disturbance. Forecasts of either variable depend upon the history of both variables. Recent efforts to arrive at workable definitions of causal orderings in economic systems have relied heavily on forecasting properties. For this reason, the forecasting equations (3) are particularly relevant for our investigation of the structure of the money stock/exchange-rate relationship.

Granger has suggested a definition of causality in which M is said to "cause" e if past values M can be used to obtain more accurate forecasts of future values of e than those forecasts formed by using only past values of e. More formally, let

$\bar{e}$ and $\overline{M}$ represent all past values of the variables e and M;

$\sigma^2(e|\bar{e})$ be the minimum predictive error variance of e given $\bar{e}$;

$\sigma^2(e|\bar{e}, \overline{M})$ be the minimum predictive error variance of e given $\bar{e}$ and $\overline{M}$.

Then, M is said to cause e if

$$(5)\quad \sigma^2(e|\bar{e}, \overline{M}) < \sigma^2(e|\bar{e})$$

Similarly, e is said to cause M if

$$(6)\quad \sigma^2(M|\overline{M}, \bar{e}) < \sigma^2(M|\overline{M})$$

When (5) and (6) both hold, feedback is said to occur. Finally, neither M nor e causes the other when

$$(7a)\quad \sigma^2(e|\bar{e}, \overline{M}) = \sigma^2(e|\bar{e})$$

$$(7b)\quad \sigma^2(M|\overline{M}, \bar{e}) = \sigma^2(M|\overline{M})$$

In this case the series are either independent or only contemporaneously related to one another. We can complete this characterization of possible causal relationships by noting that the variables are instantaneously, or contemporaneously, related if the contemporaneous value of one variable aids in forecasting the other. That is, if

$$(8a)\quad \sigma^2(e|\bar{e}, \overline{M}, M) < \sigma^2(e|\bar{e}, \overline{M})$$

The direction of instantaneous causality cannot be identified empirically; hence, (8a) implies

$$(8b)\quad \sigma^2(M|\overline{M}, \bar{e}, e) < \sigma^2(M|\overline{M}, \bar{e})$$

As has been indicated previously (see Sims, 1972, and Granger), these characterizations of causality imply specific restrictions on the parameters of the bivariate stochastic process. The restrictions of substantive importance for testing our Propositions E1–E3 and M1–M3 are indicated in the following two sections.

II. Market Efficiency Propositions and the Bivariate Stochastic Model

We now make use of the bivariate stochastic system to characterize the three market efficiency Propositions E1–E3. Fama established the convention of defining a market as efficient if market prices "fully reflect" available information. The term fully reflect conveys the notion that available information cannot be used to forecast prices in a fully efficient market because an efficient market immediately responds to

new information, rendering such information worthless as a forecasting tool. An inefficient market is one in which prices respond sluggishly to new information. In such a market a forecaster can use new information to improve forecasts of future price movements.

It must be recognized at the outset that market efficiency can be discussed only with reference to specified information sets. There are obviously some kinds of information which would be useful for forecasting prices except that such information is impossible or too costly to obtain. We consider two levels of information. The first is the past history of the price under consideration. If the past history of the price is of no value in forecasting, then the market is deemed to be *weak-form efficient*, a term first employed by Fama.

The fact that a market is weak-form efficient does not rule out the possibility that information other than the past history of the price itself might prove useful in forecasting. The second level of information therefore includes those variables that might provide useful information other than the past history of the price itself. To characterize the notion of market efficiency with respect to the second level of information, we introduce the concept of *incremental efficiency*. The market is deemed to be incrementally efficient with respect to the variables at the second information level if their use cannot improve upon the forecasts supplied by the past history of the price itself. Some emphasis must be placed upon the term incremental. It may well be that one could use variables in the second level to forecast the price series if variables in the past history of the price itself were not considered. Our convention, however, is to consider a variable as useful for forecasting only if it improves upon the forecasts yielded by the past history of the price itself.

If a market is both incrementally efficient and weak-form efficient, then there is no information at either level that is useful for forecasting. Conventional terminology refers to such a market as *semistrong-form efficient* (see Fama).

Any empirical investigation of the efficient market phenomenon must specify which variables are to be considered in the second information level. Because theoretical attention has been directed at the money stock as a primary determinant of the exchange rate, we consider the second level of information as comprising only one variable —the relative level of *U.S.* and Canadian stocks of money.

We begin our analysis with an investigation of Proposition E1—that the market is weak-form efficient. Formally, the market is weak-form efficient when

$$E(e_t)-E(e_t|\bar{e})=0 \tag{9a}$$

or, equivalently, when

$$\sigma^2(e_t|\bar{e})=\sigma^2(e_t) \tag{9b}$$

This corresponds to Fama's definition of a fair game, with e_t (the change in the exchange rate) being the fair game variable. The definitions (9a) and (9b) state that the forecast of e_t cannot be improved by conditioning on past e. If we further specify the e_t process to be stationary with a finite variance, then $E(e_t)$ will be a constant. The exchange rate then moves as a random walk, $E(e_t)=0$, or a random walk with drift, $E(e_t)=k$. The hypothesis of weak-form efficiency states that investors fully utilize past price movements in making their portfolio decisions. In so doing they cause the current price to instantly reflect all information available in past prices: therefore, past prices cannot be used to forecast price movements.

In order to test Proposition E1, we note that the change in exchange rates in our bivariate framework shown in equation (1) is represented as

$$e_t=\gamma(B)\varepsilon_{Mt}+\delta(B)\varepsilon_{et} \tag{10}$$

The parameters γ_i, δ_i, for $i=0,\ldots,\infty$, along with the error variances, σ_M^2 and σ_e^2, determine a covariance structure for e_t.

Equation (10) can be reparameterized as

$$e_t = \psi(B)u_{et} \tag{11}$$

where u_{et} is a white noise process with variance σ^2_{ue}.[6] The ψ_i's for $i=0,\ldots,\infty$, and σ^2_{ue} are functions of the γ_i, δ_i, σ^2_M, and σ^2_e, such that the covariance structure of e_t in (11) is the same as that in (10). Equation (11) represents the univariate structure for e_t implied by the bivariate system (1). Weak-form market efficiency imposes a particular restriction on (11), namely that $\psi(B)=1$, so that the exchange-rate process is a random walk.

Our empirical test of the weak-form efficiency proposition is based upon the Canadian price of *U.S.* dollars on a monthly basis, observed during the periods 1953(1)–1962(4) and 1970(7)–1975(8), when the exchange rate was allowed to fluctuate.[7] Two alternative definitions of exchange rates were used. We first employed the change in the *log* of the monthly seasonally unadjusted spot rate (e). We then repeated the empirical work adjusting spot rates for the three-month interest differential between *U.S.* and Canadian Treasury Bill rates (e^*).[8] The former series is more in the spirit of the efficient market literature and is more appropriate when testing the government reaction function. The latter series is of interest to trade theorists since it reflects the interest arbitrage conditions.

As a test for weak-form efficiency, Proposition E1, we examine the simple random walk model for spot rates in both periods. The adequacy of the random walk models was tested by calculating the statistic Q^*:

$$Q^* = N(N+2)\sum_{k=1}^{M}(N-K)^{-1}\hat{\sigma}_k^2$$

where N is the number of observations, and $\hat{\sigma}_k$ is the computed kth order autocorrelation coefficient of the residuals from the random walk model.[9] The statistic Q^* is asymptotically distributed as $\chi^2(M)$. Based upon $M=25$, we obtained Q^* statistics for e and e^* of 24.72 and 25.72 in the early period; in the later period the Q^* values were 32.63 and 33.08 for e and e^*. The $\chi^2_{(25)}$ critical value is 34.4 at the 10 percent level of significance. We are therefore unable to reject Proposition E1 and conclude that the data are consistent with weak-form efficiency in the Canadian-*U.S.* market for foreign exchange.

We now investigate Proposition E2—that the market is incrementally efficient with respect to the relative money stocks of Canada and the United States. The market is incrementally efficient with respect to the relative money stock, denoted by M, if the history of M provides no information by which the forecast of e_t conditional on its own past can be improved. Formally, incremental efficiency with respect to M is defined as

$$E(e_t|\bar{e}) - E(e_t|\bar{e},\overline{M}) = 0 \tag{12a}$$

or

$$\sigma^2(e_t|\bar{e}) = \sigma^2(e_t|\bar{e},\overline{M}) \tag{12b}$$

Stated in this form, the concept of incremental efficiency provides the appealing heuristic interpretation that the money supply cannot be used as a leading indicator

[6] See Arnold Zellner and Franz Palm.

[7] Data for the exchange rate were supplied by the Bank of Canada, Ottawa, in computer printouts. These data will be made available to readers upon request.

[8] The interest arbitrage condition implies that the fair game variable is not the exchange rate itself, but the exchange rate corrected for the interest differential. The arbitrage condition is satisfied when

$$\frac{\tilde{e}_t}{\tilde{e}_{t-1}}(1+r^{us}_{t-1}) = (1+r^c)k$$

where e is the exchange rate, and r^{us} and r^c are *U.S.* and Canadian interest rates on short-term financial instruments. The logarithmic approximation to the above equation is

$$ln\,\tilde{e}_t - ln\,\tilde{e}_{t-1} + ln(1+r^{us}_{t-1}) - ln(1+r^c_{t-1}) = ln\,k$$

or

$$e_t + (r^{us}_{t-1} - r^c_{t-1}) = k$$

Hence e^*_t is given by $e_t + r^{us}_{t-1} - r^c_{t-1}$. The correct interest rates to use are one-month yield to maturity rates. We were informed by the Bank of Canada that no such series exist for the Canadian bill market. We therefore used the three-month differential as a proxy for the one-month spread.

[9] See G. E. P. Box and G. Jenkins.

for the exchange rate. Furthermore, comparison of (12b) with (7b) reveals that our definition of incremental efficiency corresponds directly with the Granger causal statement that M does not cause e.

The equivalence of (12b) and (7b) suggests a problem in testing the incremental efficiency Proposition E2. Incremental efficiency implies that past M cannot be used to forecast current e. On the basis of (7b) however, the inability of past M to forecast current e would be interpreted as a lack of causality running from M to e. Therefore, it is not possible to distinguish between Proposition E2, that the market is incrementally efficient, and Proposition M2, that the monetary approach to exchange-rate determination is invalid. The existence of an efficient market vitiates the usual directly testable implications of the simple monetary approach to exchange-rate determination. As we shall see below, however, the possibility of a purely contemporaneous relationship between the e and M (explored in Section III) provides evidence that can be used to assess the simple monetary approach. First we proceed to explore Proposition E2, subject to the knowledge that the results of the test may be interpreted in two ways.

If Proposition E2 is true, the equation (12b) holds. This implies that $\gamma(B)$ in (1) and $c(B)$ in (3b) are equal to 0. With $c(B)=0$, (3b) reduces to

$$e_t = a(B)e_{t-1} + \varepsilon_{et} \tag{13}$$

Equation (13) shows that, when $\gamma(B)=c(B)=0$, past money does not aid in forecasting current e_t.

Sims (1972) presents a statistical procedure for testing the hypothesis that $\gamma(B)=0$. His procedure is grounded upon the fact that the bivariate stochastic process (1) yields an equation for M_t of the form

$$M_t = \sum_{j=-\infty}^{\infty} f_j e_{t-j} + \xi_t \tag{14}$$

where ξ_t is uncorrelated with the regressor matrix consisting of all future, current, and past M.[10] If the restriction $\gamma(B)=0$ is satisfied such that the money stock does not cause exchange-rate changes, then, the f_j for $j<0$ will be zero. Thus, to evaluate Proposition E2 we estimate equation (14) and jointly test the significance of the coefficients on future M.

We used two characterizations of the money stock in assessing Proposition E2. These are the growth rate of the Canadian M_1 series (M) and the difference in the growth rates between the *U.S.* and Canadian M_1 series (M^*). The Canadian M_1 series is perhaps more appropriate in examining the government reaction function whereas the difference in money growth rates corresponds more closely to both the theoretical and empirical work of authors characterizing their models as monetary approach models. The empirical results are reported for both definitions.[11] We have also employed both exchange-rate variables: e and e^* discussed above. The results were similar, so only the results based on e are reported.

Table 1 contains the results of the two-sided regressions of M and M^* on past, current, and future e. Each column of Table 1 presents the results of a regression of the dependent variable on six leading, six lagged, and the contemporaneous value of the independent variable. The estimation procedure, based on the work of Geweke, Edward Hannan, and Sims (1972), is a frequency domain procedure which is asymptotically equivalent to generalized least squares (*GLS*). Further detail is supplied in the Appendix.

To test Proposition E2, that the market is incrementally efficient, we must determine whether the six coefficients on leading values of e are equal to zero. As can be seen in Table 1, none of the individual coefficients are significantly different from zero. At the bottom of each column we present computed F-statistics, along with the 10 percent critical values, for testing the joint hypotheses that all of the future coefficients in each

[10]The ξ_t are not, however, serially independent.

[11]The data on the Canadian money supply were generously provided by the Bank of Canada and the *U.S.* M_1 series was obtained from the *Federal Reserve Bulletin.*

TABLE 1—*GLS* REGRESSION ESTIMATES OF M AND M^* ON PAST, CURRENT, AND FUTURE VALUES OF e

		Time Period			
		1953–1962		1970–1974	
		M on e	M^* on e	M on e	M^* on e
Future Coefficients	6	.302	−.325	−.544	.320
	5	−.149	.306	.391	−.367
	4	.280	−.430	.083	−.038
	3	−.050	.229	.520	−.464
	2	.094	−.272	−.042	−.090
	1	.116	.039	−.499	.428
	0	.138	−.362	.591	−.575
Past Coefficients	1	.338	−.344	−.030	.135
	2	−.280	.378	.015	.115
	3	.053	−.046	−.297	.288
	4	.233	−.158	.246	−.401
	5	−.085	.125	.392	−.462
	6	.076	−.012	.208	−.307
	R^2	.150	.208	.535	.577
Computed F		.49	.56	.70	.77
Critical Value (10 percent)		1.85	1.85	1.93	1.93

Note: M = Canadian money stock growth rate; M^* = Difference in *U.S.* and Canadian money stock growth rates; e = Growth rate in Canadian-*U.S.* exchange rate.

regression are zero. In every case the computed F is considerably below the critical level. We are therefore unable to reject the hypothesis of incremental efficiency in the Canadian-*U.S.* market for foreign exchange.

Since we find the market to be both weak-form and incrementally efficient, we conclude that none of the variables in the specified information set can be used to forecast changes in the exchange rate. Thus we are unable to reject Proposition E3 that the market displays semistrong-form efficiency. Formally, the market displays semistrong efficiency (see Fama) with respect to the information set $(\bar{e}, \overline{M})$ if

$$(15a) \qquad E(e_t) - E\left(e_t | \bar{e}, \overline{M}\right) = 0$$

or

$$(15b) \qquad \sigma^2\left(e_t | \bar{e}, \overline{M}\right) = \sigma^2(e_t)$$

The existence of semistrong efficiency implies that the entire conditioning information set is of no value in forecasting of exchange rates. The data we examined are thus fully consistent with semistrong-form efficiency in the Canadian-*U.S.* market for foreign exchange.[12]

III. The Simple Monetary Model of Exchange-Rate Determination in the Framework of the Bivariate Stochastic Process

Having completed our investigation of market efficiency, we can now turn to an examination of the stock of money in the bivariate stochastic model. The most typical assumption in the monetary approach to exchange-rate determination is that of monetary exogeneity. In our model, monetary exogeneity can be characterized by the restriction on equation (1) that

$$(16) \qquad \beta(B) = 0$$

[12]A slight modification of Barro's structural model yields some insights concerning the necessary structural conditions for semistrong efficiency. By suppressing a trend term and combining two disturbances into a single shock, Barro's money supply equation can be written in the form:

$$(a) \qquad M_t = M_{t-1} + \varepsilon_{Mt} - \theta \varepsilon_{Mt-1}$$

Sims (1972) has shown this restriction to be equivalent to the statement that the exchange rate does not cause the money supply in the framework used by Granger. Restriction (16) implies that $b(B)=0$ in (3a) and the monetary forecasting equation reduces to

$$M_t = d(B)M_{t-1} + \varepsilon_{Mt} \tag{17}$$

Equations (17) shows that, as a result of restriction (16), past exchange-rate changes do not aid in forecasting the money supply. This is precisely the restriction that must be tested in order to evaluate Proposition M1. Restriction (16) is implicitly assumed in the empirical work of Frenkel and in the studies reviewed by Magee, for only under (16) is consistent estimation possible with least squares.[13] Rather than assume monetary exogeneity with respect to exchange rates, we propose to test the restriction directly.

To assess this hypothesis we employ the Sims procedure in a manner analogous to that used to test $\gamma(B)=0$. We estimate

$$e_t = \sum_{j=-\infty}^{\infty} g_j M_{t-j} + \eta_t \tag{18}$$

and test the significance of coefficients on future M. A rejection of the hypothesis that $g_j = 0$, $j<0$, amounts to a rejection of the hypothesis that $\beta(B)=0$, and a rejection of Proposition M1.

Table 2 contains the results of the two-sided regressions of e on M and M^*. To test Proposition M1, that the money supply growth is exogenous, we must determine whether the six coefficients on leading values of M or M^* are significantly different from zero. The results are not as clear as those in Table 1. Several of the leading coefficients are individually significant. The F-statistics presented at the bottom of Table 2 are also considerably higher than the corresponding F-statistics presented in Table 1. For the early period, the F-statistics are above the critical value at the 10 percent level of significance. In the early period we find weak support for rejection of Proposition M1 that the money supply is exogenous with respect to the exchange rate. For the later period, however, Proposition M1 is not rejected. The finding that M1 is endogenous in the early period in any case raises questions about the conclusion of earlier authors, who assumed exogeneity of $M1$ in testing the simple monetary theory of exchange rates.[14]

Similarly, by omitting two-trend variables representing the foreign price level and real income, we can represent Barro's demand equation as

$$M_t = e_t - \alpha(Ee_{t+1} - e_t) \tag{b}$$

Barro's solution for e_t is given by

$$e_t = M_t - \left[\frac{\alpha\theta}{1+\alpha}\right]\varepsilon_{Mt} \tag{c}$$

$$E(e_{t+1} - e_t) = -\left(\frac{\theta}{1+\alpha}\right)\varepsilon_{Mt} \tag{d}$$

The solution equation (c) for e_t shows it to be a function of M_t which in turn by (a) can be forecast using past M. Equation (d) shows that the expectation of the change in the exchange rate is nonzero; thus, Barro's model does not represent an efficient exchange market. Since Barro's model is in *logs*, α is the elasticity of the speculative demand for money with respect to $E(e_{t+1}-e_t)$. In an efficient market α will be ∞: as long as there are expected profits (losses) from holding a currency, speculators will accumulate (sell) that currency. Speculators cease trading only when they have, in the process of adjusting their stocks, bid exchange rates to a level where expected gains and losses are zero. When the elasticity of speculative demand is infinity, we see from equation (d) that $E(e_{t+1}-e_t)=0$, as must be the case for an efficient market.

[13]If $B(B)=0$, then $M_t = \alpha(B)\varepsilon_{Mt}$ and $e_t = \gamma(B)\varepsilon_{Mt} + \delta(B)\varepsilon_{et}$. To eliminate correlation between ε_{Mt} and ε_{et} we project ε_{et} on ε_{mt}, obtaining $\varepsilon_{et} = \rho\varepsilon_{mt} + u_t$. Now, $\varepsilon_{mt} = \alpha^{-1}(B)M_t$, and $e_t = \gamma(B)\varepsilon_{mt} + \delta(B)\rho\varepsilon_{mt} + \delta(B)u_t$. So

$$\begin{aligned} e_t &= \gamma(B)\alpha^{-1}(B)M_t + \delta(B)\rho\alpha^{-1}(B)M_t + \delta(B)u_t \\ &= \alpha^{-1}(B)(\gamma(B) + \delta(B)\rho)M_t + \delta(B)u_t \\ &= k(B)M_t + h(B)u_t \end{aligned}$$

Because u_t is orthogonal to present and past ε_M, it is orthogonal to M_t, a linear combination of present and past ε_M. Therefore least squares estimates of $k(B)$ are consistent.

[14]This result is consistent with the findings of Pippenger and Phillips who "accept the hypothesis that the [Canadian] Exchange Fund played an important role ... through systematic intervention" (p. 810).

TABLE 2—*GLS* REGRESSION ESTIMATES OF e ON PAST, CURRENT, AND FUTURE VALUES OF M AND M^*

		Time Period			
		1953–1962		1970–1975	
		e on M	e on M^*	e on M	e on M^*
Future Coefficients	6	−.018	.016	.004	−.049
	5	.038	−.012	.176[a]	−.190[a]
	4	.176[a]	−.096	.119	−.134[a]
	3	−.019	.032	.033	−.032
	2	−.121	.106	−.041	.014
	1	.068	−.102	−.028	−.010
	0	.081	−.154[a]	.017	−.058
Past Coefficients	1	.036	−.071	−.001	−.034
	2	.022	−.039	.110	−.126
	3	.028	−.007	.158	−.118
	4	.079	−.047	.046	−.025
	5	−.014	.025	−.003	.003
	6	−.060	.051	−.105	.042
	R^2	.209	.221	.375	.419
Computed F		1.93[b]	1.88[b]	1.27	1.76
Critical Value (10 percent)		1.85	1.85	1.93	1.93

Note: See Table 1.
[a] indicates significance at the 5 percent level.
[b] indicates significance at the 10 percent level.

We can now direct attention to the fundamental properties of the simple monetary approach to exchange-rate determination. We are faced with having to determine whether there is any evidence that money causes exchange rates. To do so we must test Proposition M2, that money does not cause exchange rates, and Proposition M3, that the only relation between money and exchange rates is contemporaneous. If Proposition M2 holds, then $\gamma(B)=0$ in equation (1). This is exactly the restriction which was tested—and not rejected—in determining the validity of Proposition E2, that the market is incrementally efficient. It is possible then to give two interpretations to our inability to reject $\gamma(B)=0$: 1) the simple monetary approach to exchange-rate determination is valid but the market is efficient; 2) the simple monetary approach is not valid. On the basis of tests conducted thus far, it is not possible to distinguish between 1) and 2). The potential existence of an efficient market for foreign exchange has obscured the most natural set of directly testable implications of the simple monetary approach to exchange rates.

There is, however, additional evidence within the bivariate stochastic model which provides some power to distinguish between Propositions E2 and M2. The action of an efficient market does not entirely obscure the impact of money on exchange rates. In an efficient market, any unexpected change in the money supply would be reflected immediately in the exchange rate. If the monetary approach is valid we would expect to find a contemporaneous relationship between money and exchange rates. Thus, in order to test the proposition that the simple monetary approach is invalid we must rule out the existence of a contemporaneous as well as a lagged impact of money upon exchange rates.

In our representation of the bivariate stochastic process (1) we utilize the normalization restriction that $\gamma_0=B_0=0$. However it is still possible that M and e are contemporaneously related since we allow the off-diagonal element of Σ, the disturbance covariance matrix, to be nonzero. The existence of a nonzero diagonal element in Σ will show up as nonzero contemporaneous terms in equations (14) and (18). Estimates

of these coefficients are shown in Tables 1 and 2: with one exception the estimates are not significantly different from zero.

The insignificance of the contemporaneous coefficients is evidence against Proposition M3, that there is a contemporaneous relationship between money and exchange rates.[15] Coupled with our earlier failure to reject Proposition M2, the evidence suggests that the monetary model is completely inoperative. To verify this we perform the joint test that both the future and contemporaneous coefficients in (14) are zero. The F-statistics are .58 and .81 for M and M^*, respectively, in the early period; 1.02 and 1.19 for M and M^* in the later period. None of these calculated F-values exceeds the critical value at the 10 percent level of significance. These tests indicate that the data provide no statistical support for the monetary model of exchange rates. However it is interesting to note that the contemporaneous coefficient has the expected sign in every case: increases in growth rate of the Canadian money supply, absolutely or relative to the *U.S.* money supply, produce increases in the Canadian price of *U.S.* dollars. The fact that this effect is consistent with theory may constitute weak evidence that both the monetary model and the efficient markets model apply—with the latter obscuring effective detection of the former.

IV. Summary and Conclusions

We have presented a bivariate model of the determination of exchange rates which is sufficiently rich to incorporate the major features of the simple monetary approach to exchange-rate determination, the efficient markets view of foreign exchange markets, and systematic intervention effects from the monetary authorities. The model allows some refining of past concepts of market efficiency, and of the testable hypothesis of incremental efficiency. It also highlights the conceptual difficulties involved in empirical observation of the monetary mechanism in markets that may be characterized by incremental efficiency. Finally, it permits testing of the hypothesis of monetary exogeneity.

We tested six propositions relating to the money supply/exchange-rate relationship. None of the various market efficiency propositions (E1, E2, and E3) can be rejected. But Proposition M1, exogeneity of the money supply, is rejected for the earlier of our two data periods, and we also find independent support for the contention that the simple monetary approach is invalid. Thus, our finding of incremental efficiency may result simply from the fact that the money supply has no bearing on exchange rates regardless of the level of market efficiency.

The finding of causality from exchange rates to the money supply is consistent with the hypothesis of government intervention in the foreign exchange market. Since monetary endogeneity appears to be a serious possibility, our results cast suspicion on the findings of other writers whose empirical techniques require exogeneity for consistent estimation. Rather than assume exogeneity, it seems advisable to conduct exogeneity tests of the type described in Section II prior to estimating dynamic regression equations.

Our results provide yet another example of the "reverse causality" phenomenon which has appeared in other economic contexts. Although governmental intervention in exchange markets is a plausible interpretation of "reverse causality" in the present study, it should be noted that other research (see our earlier paper) has shown that such an outcome is likely to be simply a manifestation of an omitted variable problem—in this case, that efficient auction market prices may be serving as important proxy variables for the fundamental disturbances affecting the economic system. These disturbances are typically observed and acted upon by participants in efficient auction markets, but they are not observed in any systematic fashion by the analyst. Their omission from the researcher's model is what leads to the appearance of "reverse causality."

[15]A nondiagonal Σ is necessary but not sufficient for this interpretation. The finding that Σ is not diagonal indicated merely the existence of a contemporaneous relationship and not its direction. As in all statistical work, the direction of contemporaneous relationships is not identified.

An important caveat that bears mention is that the preceding conclusions are based entirely on a simple bivariate representation. Since exchange-rate models are typically quite complex in structure, the bivariate analysis may be subject to an analogue of the omitted variable problem in more conventional econometric analyses.[16] As the econometric and computational methodology for handling more complex multivariate systems becomes available, the robustness of the preceding results should be subject to further tests in the context of a multivariate specification that allows explicit modeling of some of the other arguments of the money demand function.

APPENDIX—*GLS* ESTIMATION PROCEDURE

Consider the regression model

$$\text{(A1)}\qquad dM_t = \sum_{j=-m}^{m} f_j de_{t-j} + \eta_t$$

where η_t is a random disturbance from a linear covariance stationary process with covariance matrix $\{E(\eta_t, \eta_{t-s})\} = \Omega$ for $s = 1, \ldots, T$, and T is the sample size.

We wish to obtain the Aitken estimator

$$\hat{f} = (\underline{de}'\Omega^{-1}\underline{de})^{-1}\underline{de}'\Omega^{-1}dM$$

where $\underline{de}$ is the $T \times (2m+1)$ matrix of lagged values of the explanatory variable vector de. Since Ω is unknown a priori and the inversion of a $T \times T$ matrix is computationally burdensome for moderately large samples, E. J. Hannan and Sims have proposed that the frequency domain representation of Ω be exploited to obtain the estimator,

$$\text{(A2)}\qquad \hat{\hat{f}} = (\underline{de}'\hat{\Omega}^{-1}\underline{de})^{-1}\underline{de}'\hat{\Omega}^{-1}dM$$

where

$$\text{(A3)}\qquad \hat{\Omega} = F\hat{\Sigma}F'$$

F is the $T \times T$ Fourier matrix with (j,k)th entry $T^{-1/2}e^{[2\pi i(j-1)(k-1)]/T}$ and $\hat{\Sigma}$ is a $T \times T$ diagonal matrix with diagonal elements representing equally spaced estimates of the disturbances' spectral density function ordinates. The first step in the estimation procedure is to form the estimates $\hat{\Sigma}$ from the residuals of the first pass ordinary least squares (*OLS*) regression of equation (A1). Since $\hat{\Sigma}$ is diagonal, its inversion is accomplished by taking the reciprocal of its diagonal elements. Applying the inverse operator to each side of equation (A3) yields

$$\text{(A4)}\qquad \Omega^{-1} = F\hat{\Sigma}^{-1}F'$$

since the Fourier matrix is unitary, i.e., $(FF' = I)$. The estimator $\hat{\hat{f}}$ is formed by substituting $\hat{\Omega}^{-1}$ from (A4) into (A2). The computation procedure takes advantage of the fast Fourier transform and transforms the data vector dM and de by $P = \hat{\Sigma}^{-1/2}F'$. The $\hat{\hat{f}}$ is calculated by a second pass *OLS* regression of the transformed equation (A1)

$$\text{(A1')}\qquad d\tilde{M}_t = \sum_{j=-m}^{m} f_j d\tilde{e}_{t-j} + \tilde{\eta}_t$$

where $d\tilde{M} = PdM$; $d\tilde{e} = Pde$ and $\tilde{\eta} = P\eta$

The *OLS* estimator for equation (A1′) is

$$\tilde{f} = \left(\underline{d\tilde{e}}'\,\underline{d\tilde{e}}\right)^{-1}\underline{d\tilde{e}}'\,d\tilde{M} = \hat{\hat{f}}$$

with the right-hand equality following from the equation

$$P'P = F\hat{\Sigma}^{-1/2}\hat{\Sigma}^{-1/2}F' = \hat{\Omega}^{-1}$$

[16]See Gary Skoog.

REFERENCES

R. J. Barro, "A Simple Flexible Exchange Rate Model with Uncertainty and Rational Expectations," unpublished manuscript, Univ. Rochester, Nov. 1975.

J. F. D. Bilson, "The Monetary Approach to the Exchange Rate—Some Empirical Evidence," mimeo., Northwestern Univ., Oct. 1976.

G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control,* San Francisco 1970.

D. Caves and E. Feige, "Efficient Markets, Stock Returns, the Money Supply, and the Economy: Which Tail Wags the Dog?," mimeo., Univ. Wisconsin, Apr. 1977.

R. Cooper, "Efficient Capital Markets and the Quantity Theory of Money," *J. Finance*, June 1974, *29*, 887–908.

M. Dooley and J. Shafer, "Analysis of Short-Run Exchange Rate Behavior March 1973 to September 1975," *Int. Finance Disc. Papers*, No. 76, Feb. 1976.

R. Dornbusch, "The Theory of Flexible Exchange Rate Regimes and Macroeconomic Policy," *Scand. J. Econ.*, No. 2, 1976, *78*, 255–75.

E. F. Fama, "Efficient Capital Markets: Review of Theory and Empirical Work," *J. Finance*, May 1970, *25*, 383–417.

E. Feige and R. McGee, "Money Supply Control and Lagged Reserve Accounting," *J. Money, Credit, Banking*, Nov. 1977, *9*, 536–51.

______ **and D. Pearce,** "Economically Rational Expectations: Are Innovations in the Rate of Inflation Independent of Measures in Innovations in Monetary and Fiscal Policy?," *J. Polit. Econ.*, June 1976, *84*, 499–522.

J. Frenkel, "A Monetary Approach to the Exchange Rate," *Scand. J. Econ.*, No. 2, 1976, *78*, 200–24.

H. Genberg and H. Kierzkowski, "Exchange Rates and Sectoral Problems," unpublished manuscript, Grad. Inst. Int. Stud., Geneva, Apr. 1976.

J. Geweke, "Employment Turnover and Wage Dynamics in U.S. Manufacturing," unpublished doctoral dissertation, Univ. Minnesota 1975.

______ **and E. L. Feige,** "Some Joint Tests of the Efficiency of Markets for Forward Foreign Exchange," *Rev. Econ. Statist.*, Nov. 1979, *61*.

C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrics*, July 1969, *37*, 424–38.

E. J. Hannan, "Regression for Time Series," in Murray Rosenblatt, ed., *Proceedings of a Symposium on Time Series Analysis*, New York 1963.

D. S. Kemp, "A Monetary View of the Balance of Payments," *Fed. Reserve Bank St. Louis Rev.*, Apr. 1975, *57*, 14–22.

R. E. Lucas, "Econometric Policy Evaluation: A Critique," work. paper, Carnegie-Mellon Univ. 1973.

S. Magee, "The Empirical Evidence on the Monetary Approach to the Balance of Payments and Exchange Rates," *Amer. Econ. Rev. Proc.*, May 1976, *66*, 163–70.

M. Mussa, "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy under a Regime of Controlled Floating," mimeo., Univ. Rochester 1975.

M. Parkin, "The Exchange Rate, the Balance of Payments and Monetary Policy under a Regime of Controlled Floating: A Comment," mimeo., Univ. Western Ontario 1976.

J. Pippenger, "The Canadian Experience with Flexible Exchange Rates," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 545–54.

______ **and L. Phillips,** "Stabilization of the Canadian Dollar, 1952–60," *Econometrica*, Sept. 1973, *41*, 797–815.

W. Poole, "Speculative Prices as Random Walks: An Analysis of Ten Series of Flexible Exchange Rates," *Southern Econ. J.*, Apr. 1967, *34*, 468–78.

N. P. Sargent, "Exchange Rate Flexibility and Demand for Money," presented at the American Finance Association Meetings, Atlantic City, Sept. 16–18, 1976.

T. J. Sargent, (1976a) "The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics," *J. Polit. Econ.*, June 1976, *84*, 631–40.

______, (1976b) "The Demand for Money During Hyperinflation under Rational Expectations," unpublished manuscript, Univ. Minnesota, Feb. 1976.

______ **and N. Wallace,** "Rational Expectations and the Dynamics of Hyperinflation," *Int. Econ. Rev.*, June 1973, *14*, 328–50.

C. Sims, "Money, Income, and Causality," *Amer. Econ. Rev.*, Sept. 1972, *62*, 540–52.

______, "Distributed Lags," in Michael S. Intrilligator and David A. Kendrick, eds.,

Frontiers of Quantitative Economics, New York 1974.

G. Skoog, "Causality Characterizations: Bivariate, Trivariate, and Multivariate Propositions," work. paper, Fed. Reserve Bank Minneapolis 1976.

A. Zellner and F. Palm, "Time Series Analysis and Simultaneous Equation Econometric Models," *J. Econometrics*, May 1974, *2*, 17–54.

Metzler on Classical Interest Theory

By John H. Wood*

The influence of Lloyd Metzler's famous paper, "Wealth, Saving, and the Rate of Interest," continues unabated. If anything, that influence has gathered strength in recent years. Metzler's paper has become a "must" for reading lists and collections of contributions to monetary thought. Its main argument is widely—almost generally—accepted. But most important, Metzler's framework has served as the basis for a wide variety of extensions of monetary theory that have themselves been influential in greater or lesser degrees.[1] Unlike the *Bible* or *The General Theory*, which are often cited but seldom read, it is clear that Metzler's paper is read carefully, for the framework presented in that paper is carried over without modification into a continuing stream of macro-economic analyses.

I will argue below that the influence of Metzler's paper is unwarranted. Specifically, his argument is marred by a logical inconsistency that invalidates his conclusions and many of the conclusions of models based on his framework.

Metzler's paper was intended as a criticism of the logical consistency of classical monetary theory as interpreted by A. C. Pigou (1941, 1943), Gottfried Haberler (1941), and Tibor Scitovsky. These writers had invoked the dependence of saving upon wealth as a defense against J. M. Keynes' contention that there is no reason to suppose that saving equals investment at an attainable rate of interest. That is, equilibrium may not in fact be achieved within the classical framework. These defenders of classical theory had replied that the deflation following upon an excess of intended saving over intended investment implies an increase in the real value of private money balances, and therefore an increase in private wealth. This in turn induced an increase in private consumption (a reduction in private saving) that would proceed until desired saving equaled desired investment at an attainable rate of interest.

This is where Metzler came in. He pointed out that the automatic tendency of the system towards equilibrium was only one property of classical economics. Another essential property was the determination of the rate of interest solely by real forces—by productivity and thrift. He argued that the dependence of the first property upon the saving-wealth relation invalidated the second property; that is, "...in salvaging one feature of classical economics—the automatic tendency of the system to approach a state of full employment—Pigou and Haberler have destroyed another feature, namely, the real theory of the interest rate" (1951, p. 95).

The amendment of the classical system to take account of the dependence of saving on private wealth meant, wrote Metzler, that instead of a unique equilibrium interest rate determined by "real" forces, the equilibrium rate may be influenced by monetary factors, depending upon how variations in the quantity of money are brought about. He considered two types of monetary change. "The second type of change consists of a direct increase or decrease in the money supply without any offsetting changes in private holdings of other assets" (1951, p. 97). Such a change leaves the equilibrium interest rate unaffected and is thus consistent with the classical model. A monetary change of "the first type," on the other hand, "is a change which takes place through open-market transactions of the central bank. The significant feature of this type of change is that it consists of an exchange of one form of asset for another" (1951, p. 97). Metzler argued

*Northwestern University. I am indebted to John Boyd, George Kanatas, George Borts, and an anonymous referee for many useful comments.

[1]For example, see Robert Mundell (1960, 1963, 1971), Edmund Phelps, Jerome Stein, Roger Waud, Edi Karni, Russell Boyer (1975, 1977), and William Branson and Ronald Teigen.

that monetary changes of the first type alter the equilibrium rate of interest. Thus, "In contrast to the classical doctrine, the theory of the interest rate implicit in the Scitovsky-Pigou-Haberler system is at least partly a monetary theory..." (1951, p. 94).

A few economists, most notably Haberler (1952) and Jürg Niehans (1978), have expressed misgivings about Metzler's method and conclusions. They have reminded us that Metzler's results depend upon the impact of open-market operations on the distribution of wealth between the private and government sectors. Surely, they argue, the classics never denied that redistributions of wealth affect the rate of interest.

But most workers in the field continue to accept Metzler's statement as a valid criticism of the consistency of the classical theory of interest and use his model as the foundation for their own analyses of general equilibrium systems. It has not yet been made clear however that Metzler's wealth redistributions and apparently nonclassical results were obtained by means of assumptions that are extremely nonclassical. Most important, his wealth effects were achieved by a particular form of disequilibrium trading. Classical discussions of forced saving and other circumstances in which monetary changes induce real effects took place in the context of analyses of dynamic disequilibria characterized by imperfect information, surprise, and disequilibrium trading.[2] But these characteristics of disequilibria are, by definition, excluded from general equilibria, in which there "cannot...be intrinsically a more insignificant thing...than money" (Mill, 1848, Book III, ch. 7).[3]

This does not mean that Metzler failed to make a significant contribution to monetary theory. But that contribution lies in the direction of the analysis of disequilibrium processes—it is not the first time that a paper has stimulated useful work different from that intended—and is far removed from his declared objective of demonstrating the logical inconsistency of the classical general equilibrium system.

Metzler differed from Mill not because he developed the implications of Pigou's saving-wealth relation, but because he assumed disequilibrium trading. Consequently, his analysis is not relevant to classical general equilibrium theory and is an inappropriate foundation for any general equilibrium analysis. This interpretation of Metzler is defended in Section I. Moreover, it will be shown that if disequilibrium trading is excluded so that, except for the inclusion of wealth in the savings function, Metzler's model is made consistent with classical equilibrium analysis, monetary changes of the second type not only leave the equilibrium rate of interest unaltered but the price level is undetermined.

Mundell (1960) improved Metzler's analysis by taking explicit account of the government's budget restraint, specifically by stating precisely the relations between tax changes and changes in government interest payments caused by variations in private ownership of government securities due to open-market operations. But the basic defects of Metzler's model were left untouched. Disequilibrium trading was retained by Mundell and the other properties of his model, like Metzler's, were such that the price level became indeterminate if equilibrium trading were imposed. These assertions are supported in Section II.

Section III presents an attempt to correct the deficiencies of the Metzler-Mundell critique of classical theory. Equilibrium trading is imposed on their system and the central bank is required to conduct open-market operations in securities representing claims on payments fixed in nominal terms (a condition permitted by neither Metzler nor Mundell) so that equilibrium trading becomes consistent with a determinate price level. It is shown that open-market operations induce alterations in the equilibrium interest rate only if they are accompanied by a change in the tax rate on personal income, which alters the rate of transformation of current saving into future consumption. This result is perfectly consistent with

[2]For example, see Henry Thornton (ch. 10), John Stuart Mill (1844), and the references in Niehans (1978, p. 86).

[3]See J. R. Hicks (1967, ch. 9) for excellent discussions of both dynamics and static equilibria in classical economics.

classical analysis and is not dependent upon the saving-wealth relation. The criticism of classical theory of Metzler and his followers thus falls to the ground.

I

Metzler wrote his system as follows (1951, p. 113):

$$S(r,w)=I(r) \tag{1}$$

$$L(r)=\frac{m}{\lambda a} \tag{2}$$

$$w=\lambda a+m \tag{3}$$

$$a=\frac{cy}{r} \tag{4}$$

where r, w, m, and a are the unknowns, r is the rate of interest (yield on common stock), w is real private wealth, m is real private money holdings, a is the real value of common stock, λ is the proportion of common stock that is privately held, y is real national income, c is the proportion of income that goes to business profits, and S and I are real current saving and investment, respectively.

Equation (1) sets saving, which is a function of private wealth and the rate of interest, equal to investment, which depends only on the rate of interest; (2) indicates the proportion in which asset holders wish to hold money and other forms of wealth; (3) defines real private wealth as the sum of privately held real money balances and common stock; and (4) results from the assumptions that real income is fixed and profits are a constant proportion of income so that the value of common stock is the capitalized value of profits.

This system may be reduced to the following two equilibrium conditions by substituting (3) and (4) into (1) and (2), rearranging, and writing the nominal money stock M and the price level P explicitly:

$$S\left(r,\frac{\lambda cy}{r}+\frac{M}{P}\right)=I(r) \tag{5}$$

$$L(r)\frac{\lambda cy}{r}=\frac{M}{P} \tag{6}$$

Equations (5) and (6) contain two unknowns, r and P.

Now consider a monetary change of the second type. This is one in which, "while other things initially remain unchanged, the quantity of money is arbitrarily doubled by giving to each holder of money an additional quantity equal to the amount he already holds" (Metzler, 1951, p. 106). This is the outside-money economy that Don Patinkin (1965) has made so familiar. Clearly, given fixed λ, c, and y, and assuming the uniqueness and stability of equilibrium, a doubling of M causes a doubling of P with no effect on r. Money is neutral, as in the classical system.

Much has been written about the conditions necessary to the uniqueness and stability of equilibrium in monetary systems. Most of that literature lies outside the scope of the present paper but I want to draw attention to one widely recognized condition necessary to uniqueness, namely, the exclusion of trading at disequilibrium prices or, using Hicks' terminology, false trading. False trading is usually avoided by making provision for recontracting. This prevents the execution of trades agreed at disequilibrium prices and assures that the final equilibrium set of prices is unique and unaffected by the precise dynamic route by which the system reaches equilibrium.[4] Hicks pointed out that false trading "has the same sort of effect as a redistribution of wealth" (1946, p. 128). Robert Clower has argued that the essential difference between Keynesian and classical economics is that the former allows disequilibrium trading while the latter does not. But all this is familiar. Suffice it to say that disequilibrium trading must not occur if we are to realize the classical neutrality of money in Metzler's monetary change of the second type.

Now let us go to a monetary change of the first type.

The effect of open-market transactions upon the equilibrium of the system

[4]See Don Patinkin (1965, pp. 40, 532) for discussions of Walrasian tâtonnement and the importance of equilibrium trading to the uniqueness of equilibrium.

> can be described in terms of a ratio indicating the proportion of the total supply of securities held in private hands [λ].... Consider, first, the situation in which λ has a value of 1.0. This means that the total available supply of securities is held by private assetholders, so that the central bank's assets consist exclusively of currency.
> [Metzler, 1951, p. 107]

Metzler then differentiates (1)–(4) with respect to λ, which is equivalent to differentiating (5)–(6) with respect to λ where m is replaced by M/P. We see in (5) and (6) that when λ is changed, r and P are affected. Apparently, the classical neutrality of money breaks down in the presence of a monetary change of the first type.[5]

But doubts appear when questions are raised concerning the direct effects of λ on w and M, some of which are not immediately apparent from Metzler's statement of his system. Once we begin to inspect the nature of the money creation process in Metzler's model, we find that he has understated the differences between his two types of monetary change. Metzler is not, in fact, distinguishing the effects of two kinds of disturbance within an unvarying framework. Rather, he deals successively with two completely different monetary systems. Specifically, his monetary changes of the first and second types correspond to inside- and outside-money systems, respectively. These terms are used in the sense of John Gurley and Edward Shaw, who defined outside money as

> ...government debt issued in payment for goods and services or in transfer payments. It [is] a claim held by consumers and firms against government. From the standpoint of the private sectors, it [is] a net external or outside claim. [Inside money, on the other hand]...is still government debt, but it is issued in payment for government purchases of private securities. It is a claim of consumers and firms against the world outside the private sectors, but it is counterbalanced by private debt to the world outside, that is, to government.... [pp. 72–73]

Clearly, Metzler's monetary change of the second type assumes an outside-money system, for which the definition of private wealth (3) and the assumption of an exogenous money supply are appropriate. But his monetary change of the first type presupposes an inside-money system and implies a different specification of wealth. It also requires an additional equation to determine the nominal money stock.

What is the private sector's wealth in an inside-money system? As in any closed system, we can ascertain a sector's wealth (net worth) only by inspecting the balance sheets of all sectors. The private sector's balance sheet, in nominal terms, is

$$(7) \quad M+\lambda A=W \quad \text{where } A=Pa;\ W=Pw$$

The central bank's balance sheet is

$$(8) \qquad (1-\lambda)A=M$$

This reflects the fact that money (bank liabilities) is injected into the system by security purchases. These two balance sheets are implied by Metzler's discussion, the first by his equation (3) and the second by the last quotation from Metzler above. Substituting (8) into (7) gives

$$(9) \quad W=M+\lambda A=(1-\lambda)A+\lambda A=A$$

It can be seen from (7) that the private sector's wealth consists of its money and security holdings. Metzler stops with this statement, which is irreproachable—as far as it goes. But then we see in (9) that private wealth is equal to the total value of securities. These two statements are reconciled by

[5] An open-market purchase by the central bank transfers the income on securities from the former owners to the bank. In order to avoid the complications introduced by this effect of monetary policy on disposable income (and to maintain the government's budget in balance), Metzler assumed that the "...profits which the central bank earns by reason of its acquisition of securities are ultimately passed on to private hands in the form of reduced taxes. Under these circumstances, the security purchases by the bank will redistribute income between former asset owners and taxpayers but will not influence the total of disposable income" (1951, p. 109, fn. 15).

(8), which shows that the private sector acquires money in exchange for securities.

We now have in (9) a specification of private wealth appropriate to an inside-money economy. We also have in (8) the equation needed to determine M when M is inside money. Rearranging, the money-supply equation may be written

$$\text{(10)} \qquad \frac{M}{P}=(1-\lambda)a \quad \text{or} \quad M=(1-\lambda)aP$$

Substituting the specification of wealth from (9) and the supply of money from (10) into Metzler's (1) and (2) gives

$$\text{(11)} \qquad S\left(r,\frac{cy}{r}\right)=I(r)$$

$$\text{(12)} \qquad L(r)=\frac{(1-\lambda)}{\lambda}$$

We see from (11) that the equilibrium rate of interest is determined, as in the classical system, in the commodity market in contradiction to Metzler's main thesis. This result differs from Metzler's because we have required that the central bank and private wealth holders exchange currency and securities only at equilibrium prices. A consequence of this requirement is that private wealth is independent of the central bank's security transactions. Metzler's specification of private wealth, on the other hand, follows from the assumption that the exchange takes place entirely at the price existing in the preceding, predisturbance equilibrium.[6]

His distinction between the effects of monetary changes of the first and second types thus disappears. If equilibrium trading is required, both types of change are neutral; if disequilibrium trading is allowed, both have real effects.

There is, of course, nothing inherently wrong with models in which disequilibrium trading is permitted. A story in which central bank security transactions are executed, all in a block, at the previously existing equilibrium price may be no further from reality than one in which those transactions take place, all in a block, at the new equilibrium price. Harry Johnson has written that "...subsequent writers, including Patinkin, have accepted this as a legitimate assumption, and Gurley and Shaw's analysis builds on it" (p. 341). Johnson was wrong about Patinkin and Gurley and Shaw, who required that all exchanges take place at equilibrium prices.[7] But his comment is applicable to several others.[8] Either approach

[6]Metzler's results (especially see his appendix) may be obtained by assuming that all security-currency exchanges occur at the predisturbance set (P_0, r_0) so that equilibrium $M=(1-\lambda)P_0cy/r_0$. (The price level is determined in Metzler's disequilibrium-trading system because the equilibrium price level P impinges on the equilibrium real money stock m.) Of course, Metzler's results, or any results, are consistent with an infinity of assumptions about the prices at which securities are traded. Metzler himself assumed that "securities will be purchased at many different prices...[But he believed that]... In view of our interest in the equilibrium of the system, we may pass over the dynamic problems and investigate, instead, the central bank's security purchases upon...the ultimate resting places of our variables" (1951, p. 108). I have argued that (a) these problems cannot be passed over if we wish to know what those ultimate resting places will be, and (b) the classical results require a particular solution to these problems, one very different from that implicitly chosen by Metzler. See George Horwich (1962, 1964) for the impacts within a dynamic framework of alternative assumptions regarding the prices at which securities are traded.

[7]Patinkin's (1965, pp. 295–305; 1969) insistence that open-market operations leave private wealth unaffected in an inside-money economy is identical to John Gurley and Edward Shaw's analysis (pp. 68–75, 140–49) in this respect. In both works the nonneutrality of money depends upon the existence of a hybrid inside-outside-money system in which monetary disturbances of either type alter the equilibrium interest rate.

[8]Branson and Teigen also implicitly assumed disequilibrium trading when they differentiated (5)–(6) with respect to M and λ, and then set $dM=-(Pcy/r)d\lambda$. This statement is true only when P and r are assumed fixed or if the P and r appearing in this equation are really meant to be P_0 and r_0. Branson and Teigen's solution is identical to that obtained from Metzler's disequilibrium trading model and is therefore consistent with the latter assumption. If they had wished to impose equilibrium trading they would have had to write $dM=-(Pcy/r)d\lambda+mdP-(M/r)dr$ and would have found two solutions for dr (one each for (11) and (12)) and none for dP. This criticism of Metzler and of Branson and Teigen is similar to Niehans' (1976) comment on the analysis presented by Alan Blinder and Robert Solow. Mundell (1960), Phelps, Stein, and Waud also unquestioningly accepted

to trading and the accompanying specification of wealth may, depending upon the objectives of the analysis, be appropriate. But one should not imagine, if the Metzlerian approach is adopted, that the results so obtained yield any insights into the logical structure of the classical model.

Moving to (12), we see that, as in (11), r is the sole unknown; P does not appear. That is, the imposition of equilibrium trading on Metzler's system leaves the price level undetermined and causes the rate of interest to be overdetermined. This is because Metzler has made his money supply proportional to the nominal value of securities, which is in turn proportional to nominal income. Metzler's system is identical in these respects to one of the inside-money models discussed by Patinkin: that in which the banking system extends credit at some arbitrary rate of interest "in accordance with the 'legitimate needs of business'—where these needs are measured in money terms, and thus increase proportionately with the price level" (Patinkin, 1965, p. 309). We see from (10) that this describes the money supply process implied by Metzler's remaining assumptions when disequilibrium trading is proscribed. The "natural" rate of interest, that is, the r that equilibrates the commodity market, is independent of open-market operations. However, this natural rate is unlikely to be achieved as long as the money supply is determined according to (10).

Price indeterminacy is encountered in Metzler's inside-money regime, when the specifications of money and wealth are made consistent with equilibrium trading, because of the choice of common stock as the security traded in his model:

> Common stock has been selected as the typical security in order to avoid the difficulties associated with bonds during periods of inflation or deflation. Throughout the paper I assume that, in the absence of movements in interest rates, common stock prices rise or fall to the same extent that other prices rise or fall, so that a general inflation or deflation does not affect the real value of securities.
> [1951, p. 99, fn. 11]

This rationale is appealing. Common stock has much to recommend its adoption as the "typical security" in macro-economic analysis. And in many models, such as the outside-money system considered in Metzler's discussion of a monetary change of the second type, the use of common stock causes no problems. But in the inside-money economy implied by a monetary change of the first type, the very "difficulties" that Metzler wished to avoid might have, if not avoided, rendered the price level determinate. What is needed for the determinacy of the price level is the impingement of the price level upon the real supply of money. It will be shown (not for the first time) in Section III that this effect may be achieved by open-market operations in securities representing claims on payments fixed in nominal terms.

II

It was shown in footnote 5 that a crucial assumption in Metzler's analysis was that income earned by the central bank on its security holdings was returned "to private hands in the form of reduced taxes." Then, on the basis of the procedure described in Section I, Metzler concluded that open-market purchases reduce the rate of interest. But Mundell (1960) argued that the "method by which taxes are reduced is crucial to the final result." He agreed with Metzler's results only if the fall in government interest payments is accompanied by a reduction in personal income taxes. "If income taxes are reduced and the capital market is not so perfect that it can discount any certain income stream, the rate of interest is lowered by open-market purchases. But if corporation taxes are lowered, the rate of interest rises" (1960, p. 624).

Metzler's redistribution of wealth through open-market operations in ostensibly equilibrium-trading models. On the other hand, Mundell (1963), Karni, and Boyer (1975, 1977) limited their use of Metzler's framework to the outside-money case and thus did not become involved with disequilibrium trading.

Mundell developed this argument within a framework that differed from Metzler's in two respects: First, he introduced government bonds—assumed to be perfectly substitutable for common stock—and let the central bank conduct open-market operations only in government debt. Second, he allowed for the possibility that the demand for money might be nonhomogeneous in other forms of wealth—this may be seen by comparing (14) with (2) and (6). Mundell wrote his system without taxes as[9]

$$S\left(r, \frac{M}{P} + \frac{cy}{r} + \frac{\lambda b}{r}\right) = I(r) \tag{13}$$

$$\frac{cy}{r} + \frac{\lambda b}{r} = B\left(r, \frac{M}{P}\right) \tag{14}$$

where b is the (fixed) real value of government interest payments, λ is now the proportion of *government* securities owned by the private sector, and (14) defines the desired relation between money and other forms of private wealth.

Now consider taxes. If the tax rate on corporate profits is γ and the government's budget is kept in balance by equating corporate taxes to interest payments, i.e., $\gamma cy = \lambda b$, the private sector's net income from securities is $(1-\gamma)cy + \lambda b = (1-\gamma)cy + \gamma cy = cy$ and (13)–(14) may be written

$$S\left(r, \frac{M}{P} + \frac{cy}{r}\right) = I(r) \tag{15}$$

$$\frac{cy}{r} = B\left(r, \frac{M}{P}\right) \tag{16}$$

Under these conditions open-market operations have no effect on private wealth. As far as we can see from (15)–(16), a change in M, due to open-market operations or not, induces an equiproportional change in P and leaves r unaffected. But at this point Mundell argued that the reduction in corporate taxes increases the attractiveness of previously undesirable projects: "Investment thus increases..., inducing a higher rate of interest" (1960, p. 625). That is, Mundell *really* meant the demand for investment goods to take the form $I(r,\gamma)$, where $\partial I/\partial\gamma < 0$.

Mundell's interesting paper increased our understanding of Metzler's model. Unfortunately, however, it brought us no closer to an understanding of the implications of the saving-wealth relation for the impact of open-market operations on the rate of interest in the classical equilibrium model. The remainder of this section is devoted first to stating Mundell's framework in a more explicit, detailed form than he gave us, and then to a demonstration that Mundell's system suffers from the same defects as those pointed out in Section I regarding Metzler's model. In particular, (i) the results emphasized by both authors depend upon disequilibrium trading, and (ii) the imposition of equilibrium trading leaves the price level undetermined and the rate of interest overdetermined.

The full government budget restraint implied by Mundell's discussion may be written as

$$\gamma cy + \rho(1-\gamma)cy + \rho\gamma b + \rho(1-c)y = \lambda b \tag{17}$$

where ρ is the personal tax rate and the terms on the left side of (17) are government receipts in real terms due to taxes on corporate profits, dividends, private interest earnings on government securities, and wages. The sum of these receipts equals government interest payments, which are assumed to be the only form of government outlay.

Real private wealth is

$$w = \frac{M}{P} + \frac{(1-\rho)(1-\gamma)cy}{r} + \frac{(1-\rho)\lambda b}{r} + \frac{(1-\rho)(1-c)y}{r} \tag{18}$$

where w is the sum of real money balances and capitalized after-tax dividends, interest, and wages. The difference between (18) and the Metzler-Mundell definitions is that capitalized wages are explicitly counted as part of private wealth. We follow Irving Fisher in allowing the anticipation of all forms of future income.

[9] Mundell's notation has been altered to maintain consistency with Metzler's presentation without, I think, affecting his argument.

The equilibrium nominal value of government securities is their capitalized after-tax nominal coupon payments, $(1-\rho)Pb/r$. The proportion of these securities owned by the central bank, obtained in exchange for currency (money), is $(1-\lambda)$. Deflating by the equilibrium commodity price level gives the equilibrium real value of central bank bond holdings, and therefore real money balances.

$$\frac{M}{P}=\frac{(1-\lambda)(1-\rho)b}{r} \tag{19}$$

In the disequilibrium trading regime of Metzler and Mundell, equilibrium real balances are

$$\frac{M_0}{P}=\frac{(1-\lambda)(1-\rho_0)P_0b}{Pr_0} \tag{20}$$

In this case, the nominal value of securities owned by the central bank is determined by the price level P_0, interest rate r_0, and personal tax rate ρ_0 existing prior to the open-market disturbance. As indicated in footnote 6, Metzler's results are consistent with this kind of equilibrium trading. This is also true of Mundell's results.

Substituting (17) into (18) gives

$$w=\frac{M}{P}+\frac{y}{r} \tag{21}$$

with equilibrium trading or

$$w=\frac{M_0}{P}+\frac{y}{r} \tag{22}$$

with Metzler-Mundell disequilibrium trading.

We see that in either regime inside money is now a net component of wealth. When the government in the exercise of its central banking function exchanges currency for some quantity of bonds valued at, say $[(1-\rho)Pb/r]d\lambda$, private wealth is not affected. But then, in order to maintain a balanced budget, the government reduces taxes by an amount equal to the reduction in interest payments to the public, that is, by $(1-\rho)Pb\,d\lambda$ with a capitalized value $[(1-\rho)Pb/r]d\lambda$ equal to the increase in the money stock.[10]

This result agrees with both Patinkin (1965, p. 307) and the controversial argument by Boris Pesek and Thomas Saving (1967, p. 220) that all money, including inside money, represents a net contribution to private wealth. The agreement exists, however, only in the absence of a private banking sector. Allowing for banks and assuming —with Patinkin but contrary to Pesek and Saving—that deposits in private banks are debts of those banks, wealth in an inside-money, balanced government budget system is

$$w=\frac{M}{hP}+\frac{y}{r}=\frac{H}{P}+\frac{y}{r}=\frac{(1-\lambda)(1-\rho)b}{r}+\frac{y}{r} \tag{23}$$

where H is central bank credit (high-powered money) and h is the money multiplier. Equation (23) applies to the case of equilibrium trading; in the event of disequilibrium trading, $H/P=(1-\lambda)(1-\rho_0)P_0b/Pr_0$.

Now let us substitute (19) and (21) into Mundell's system (13)–(14) and see how his analysis holds up under classical assumptions, including perfect capital markets and equilibrium trading. Commodity market equilibrium may be written as

$$S\left(r,\frac{(1-\lambda)(1-\rho)b}{r}+\frac{y}{r}\right)=I(r,\gamma) \tag{24}$$

[10]The difference between (21) and the definition of $w(=cy/r)$ in (11) is due to Metzler's assumptions of no government bonds or profits taxes in combination with his assumption that only wage income is taxed (or subsidized). His goverment budget restraint is, apparently, $(1-\lambda)cy+\rho(1-c)y=0$ so that $\rho=-(1-\lambda)c/(1-c)$, where ρ is a subsidy rather than a tax. Ignoring human wealth, $w=m+\lambda cy/r=(1-\lambda)cy/r+\lambda cy/r=cy/r$, where ρ affects neither m nor $\lambda cy/r$. But if human wealth is added to w and the above solution of ρ from the government budget restraint is substituted into the result, we get $w=cy/r+(1-\rho)(1-c)y/r=y/r+(1-\lambda)cy/r=y/r+m$.

Consistent with our assumption that all forms of wealth are equally important to consumption and portfolio decisions, human wealth should be added to the left-hand side of (14) and the result may be transformed to obtain the usual demand for money statement: $M/P = L(r,w)$; or, using (19) and (21),

$$(25)\quad \frac{(1-\lambda)(1-\rho)b}{r} = L\left(r, \frac{(1-\lambda)(1-\rho)b}{r} + \frac{y}{r}\right)$$

It can be seen from (24)–(25) that, just as when equilibrium trading was imposed on the Metzler model in Section I, the price level is undetermined. This is because Mundell has fixed government interest payments in real terms. His bonds are of the constant purchasing power variety. Consequently, when the central bank attempts to add to (or subtract from) its security holdings, the price level is bid up (or down) without limit, for there are no forces to keep P in check and to constrain it to move towards an equilibrium.

As in Metzler's system, determinacy of the price level may be achieved if disequilibrium trading is assumed. Mundell in fact followed Metzler in adopting a trading regime in which all open-market transactions take place at the predisturbance price of securities.[11] This means that Mundell's system, altered by our assumption of perfect capital markets, is obtained by using (20) and (22) instead of (19) and (21). That is, the final equilibrium interest rate and price level in the Metzler-Mundell disequilibrium trading regime are determined by (26)–(27) if capital markets are perfect.

$$(26)\quad S\left(r, \frac{(1-\lambda)(1-\rho_0)P_0 b}{Pr_0} + \frac{y}{r}\right) = I(r,\gamma)$$

$$(27)\quad \frac{(1-\lambda)(1-\rho_0)P_0 b}{Pr_0} = L\left(r, \frac{(1-\lambda)(1-\rho_0)P_0 b}{Pr_0} + \frac{y}{r}\right)$$

Consistent with Mundell's main argument, the impact of open-market operations on the rate of interest in this system depends upon the means by which the government's budget restraint is satisfied, that is, upon whether the corporate or the personal income tax rate is altered to compensate for changes in government interest payments. For example, a doubling of the proportion of government securities held by the central bank $(1-\lambda)$, if accompanied by a change in the personal tax rate ρ and an unchanged corporate tax rate γ, will cause a doubling of the equilibrium price level P and leave the real sector unaffected.[12] However, if ρ is kept constant and the government's restraint is satisfied by changes in γ, then

[11]"...[O]pen-market operations...involve an exchange of money and government securities between the central bank (or government) and the public. But, since the money component of this exchange does not affect the final equilibrium, open-market operations can be analyzed as a single change in outstanding government debt" (Mundell, 1960, p. 623).

[12]Remember that ρ_0 is given from the predisturbance situation. The government's budget is therefore kept in balance by altering the personal tax rate on current income, which means changes in ρ, not in ρ_0, which is the rate existing when Metzler-Mundell open-market operations are carried out. The Metzler-Mundell imperfect-capital-market results are obtained when we deduct the capitalized human income stream $(1-\rho)(1-c)y/r$ from the wealth terms in (26)–(27); i.e., from the second arguments in $S(\cdot)$ and $L(\cdot)$. Under these conditions, an open-market purchase will, when accompanied by a reduction in ρ, cause a decline in the equilibrium rate of interest. Pesek and Saving (1963) have set forth a model that they represent as a continuation of the Pigou-Haberler-Scitovsky-Metzler discussion and in which the rate of interest is a "fully monetary" phenomenon; i.e., *every* kind of monetary change affects the equilibrium rate of interest. They assume equilibrium trading and their results depend on a combination of imperfect capital markets and aggregate tax bills fixed in nominal terms.

investment demand, and therefore r, will be altered by open-market operations.[13]

III

My main concern is not with the Metzler-Mundell disequilibrium trading model. Rather, my objective is identical to that declared, but not pursued, by Metzler—to determine the influence of open-market operations on the interest rate in a general equilibrium system that differs from classical statements only by the addition of the saving-wealth relation.[14] We saw in Section I that this objective cannot be attained within Metzler's framework because of the perfectly elastic money supply implied by making Metzler's model conform to classical assumptions, a course that led to an undetermined price level and an overdetermined rate of interest. We saw further in Section II that Mundell's model suffers from the same deficiencies. I can see no way around this problem where Metzler's model is concerned; any inside-money model in which common stock is the only non-monetary asset must, under classical assumptions, result in an undetermined P and the overdetermination of r. But there is a way out in Mundell's case. All that is needed to achieve price determinacy and a unique solution for r when Mundell's system is forced to satisfy classical assumptions is to let government interest payments be fixed in nominal terms instead of being linked to the price level. Then attempts by the central bank to acquire a greater share of government debt will not be defeated by the unlimited increases in nominal interest payments required to maintain the real value of those payments. This alteration in Mundell's model is recognized in (28)–(29), where b has been replaced by B/P and now the nominal value of government interest payments B is fixed.

But we are not yet ready to determine the impact of open-market operations on the rate of interest in the classical general equilibrium model with wealth effects. This is because the system described by (24)–(25) does not properly express the influence of income taxes on spending and portfolio decisions. The basic error along these lines in the Metzler-Mundell model, which has been carried on by other writers, was their failure to distinguish the before-tax rate of interest from the after-tax rate of return. Specifically, they failed to recognize that r in their models is the after-tax rate of return, which is related to the market rate of interest (defined hereafter as i) as follows: $r = i(1-\rho)$. In a regime in which interest income is taxable and interest costs are tax deductible at the rate ρ, the cost of current consumption (the return to saving) is $[1+i(1-\rho)] = (1+r)$. This is also the appropriate discount factor to be applied to future income in the valuation of wealth.[15] It is also the effective

[13]These results are identical to Metzler's 1973 conclusions regarding the perfect capital market case in his response to Mundell's extension of Metzler's 1951 paper—except that Metzler felt corporate profits taxes to be so small relative to other taxes that they may safely be neglected. Hence, "when the capital market is perfect according to Fisher's definition, the rate of interest remains unchanged by open-market operations" (Metzler, 1973, p. 355). But we will see in Section III that these results were due partly to Metzler's and Mundell's neglect of the impact of personal income taxes on saving-consumption decisions.

[14]This objective might reasonably be understood to imply a strict Cambridge or Fisher transactions demand for money, for example, $M/P = ky$ instead of the liquidity preference function (25). It is shown in the Appendix that the qualitative results obtained below are the same for both of these money demand functions.

[15]Consider, for example, the two-period case in which a consumer wishes to maximize $U(c_1, c_2)$ subject to the budget restraint

$$c_2 = [y_1(1-\rho) - c_1](1+i) - \rho i[y_1(1-\rho) - c_1] + y_2(1-\rho)$$
$$= [y_1(1-\rho) - c_1][1+i(1-\rho)] + y_2(1-\rho)$$

where y_t and c_t are income and consumption in the t-th period and $\rho i[y_1(1-\rho)-c_1]$ is the tax or tax deduction on interest earnings or payments. The consumer will arrange consumption such that $U_1/U_2 = [1+i(1-\rho)]$, where U_t is the marginal utility of consumption in the t-th period. The budget restraint may be rewritten

$$c_1 = \frac{c_2}{[1+i(1-\rho)]} = y_1(1-\rho) + \frac{y_2(1-\rho)}{[1+i(1-\rho)]}$$

where the right-hand side of the equation is the consumer's wealth, i.e., the present value of his income stream. In a steady state with an infinite horizon, the present value of future income is $y(1-\rho)/i(1-\rho)$ before taking account of the government's budget restraint and $y/i(1-\rho) = y/r$ after taking account of that restraint.

(after-tax) cost of holding money rather than securities. Consequently, r in (24)–(25) may everywhere be replaced by $i(1-\rho)$. This has been done on the right-hand sides of (28)–(29), except in the case of the demand for investment goods. Joseph Stiglitz (1973, 1976) has shown that, in a regime such as ours, with perfect capital markets, investment by firms is independent of corporate profits taxes. That investment plan which maximizes before-tax earnings also maximizes after-tax earnings.[16]

These corrections mean that (24)–(25) must be rewritten as follows, where the left-hand and right-hand statements in r and i, respectively, are equivalent:

(28)

$$S\left(r, \frac{(1-\lambda)(1-\rho)B}{Pr} + \frac{y}{r}\right) - I\left(\frac{r}{1-\rho}\right) = 0$$

$$= S\left(i(1-\rho), \frac{(1-\lambda)B}{Pi} + \frac{y}{i(1-\rho)}\right) - I(i)$$

(29)

$$L\left(r, \frac{(1-\lambda)(1-\rho)B}{Pr} + \frac{y}{r}\right) - \frac{(1-\lambda)(1-\rho)B}{Pr} = 0$$

$$= L\left(i(1-\rho), \frac{(1-\lambda)B}{Pi} + \frac{y}{i(1-\rho)}\right) - \frac{(1-\lambda)B}{Pi}$$

It will be convenient to rearrange the government budget restraint (17) to make clear the dependence of $(1-\rho)$ upon $(1-\lambda)$ and P:

(30)

$$(1-\rho) = \frac{y}{(1-\gamma c)y + B/P - (1-\lambda)B/P}$$

Clearly, open-market operations are neutral in this system if the personal tax rate is held fixed, that is, if the corporate profits tax γ is altered in response to variations in λ and ρ such that the government's budget is kept in balance. Under these conditions, a doubling of the central bank's share of ownership of government debt $(1-\lambda)$ induces a doubling of the price level with no effect on r or i.

Open-market operations are not neutral however when accompanied by alterations in the personal tax rate.[17] This may be shown as follows: Suppose an open-market purchase—a doubling of $(1-\lambda)$, for example—induces a doubling of the price level and leaves the rate of interest i unchanged. This is not consistent with equilibrium if the personal tax rate has been lowered to offset the fall in the government's interest payments λB to the public. We see from the right-hand side of (28) that an increase in $(1-\rho)$ disturbs both the substitution and wealth arguments in the savings function. The effective return to saving is increased and the tax-adjusted capitalized income stream is reduced. Each of these factors acts to increase saving, and therefore to reduce i. The decline in i is reinforced by the decrease in the demand for money and resulting further increase in the price level due to the influence of $(1-\rho)$ on the substitution and wealth terms shown on the right-hand side of (29). The extent of the fall in the equilibrium i due to an open-market purchase is shown in equation (A2) in the Appendix.

The equilibrium after-tax rate of return r is raised by an open-market purchase that is accompanied by a reduction in the personal tax rate. This effect is caused by the increase in the demand for investment goods due to the decline in i. We see from the left-hand sides of (28) and (29) that an open-market purchase would leave r unaffected and cause a more than proportional increase in the price level if investment were interest inelastic.[18] But if $dI/di<0$, where

[16]Stiglitz has thus generalized Knut Wicksell's one period result, in which production decisions by profit-maximizing firms are independent of profits taxes.

[17]Metzler's (1973) contention that personal income taxes dominate corporate taxes (see fn. 13) is thus seen to produce results just the opposite of his.

[18]The change in the price level that is consistent with equilibrium when investment is interest inelastic and $(1-\rho)$ is dependent on $(1-\lambda)$ and P as shown in (30)

$i=r/(1-\rho)$, the increase in $(1-\rho)$ induces an upward shift in investment demand and, therefore, an increase in r. The positive influence of an open-market purchase on the equilibrium value of r is shown precisely in equation (A1).

In summary, the wealth-saving relation is neither a necessary nor a sufficient condition for the nonneutrality of open-market operations in a general equilibrium system. The key factor is fiscal policy. Whether or not open-market operations induce changes in either the before-tax market rate of interest i, or the after-tax rate of return r, depends upon the fiscal actions chosen to accompany the actions of the central bank. So in the end we arrive at the almost inevitable conclusion—that classical assumptions produce classical results. For, surely, a change in the equilibrium rate of interest because of an alteration in the rate of transformation of current saving into future consumption—due to a change in the personal tax rate, as in my model, or for any other reason—is consistent with, even a requirement of, classical analysis.

Appendix

This Appendix is concerned with the comparative statics of equations (28)–(29), which make up the only system in this paper that is characterized by both equilibrium trading and a determinate price level. In particular, we wish to know the effects of open-market operations on i, r, and P when the corporate tax rate γ is held fixed so that the government's budget restraint must be satisfied by adjustments in the personal tax rate ρ. These effects are found by totally differentiating (28)–(29) with respect to $(1-\lambda)$ while taking account of the dependence of ρ on λ and P shown in (30):

$$\text{(A1)} \quad \frac{dr}{d(1-\lambda)} = -\frac{B(E_\lambda + E_p)(1-L_w)I'}{P[(1-E_p)mG - E_pK]}$$

$$\text{(A2)} \quad \frac{di}{d(1-\lambda)} = -\frac{B(E_\lambda+E_p)[(1-L_w)S_r + S_w(L_r - y/r^2)]}{P[(1-E_p)mG-E_pK]}$$

$$\text{(A3)} \quad \frac{dP}{d(1-\lambda)} = \frac{P[(1+E_\lambda)mG + E_\lambda K]}{(1-\lambda)[(1-E_p)mG - E_pK]}$$

where

$$S_r = \frac{\partial S}{\partial r}; \quad S_w = \frac{\partial S}{\partial w}; \quad I' = \frac{\partial I}{\partial i}$$

$$L_r = \frac{\partial L}{\partial r}; \quad L_w = \frac{\partial L}{\partial w}$$

$$G = \left[(1-L_w)\left(S_r - \frac{I'}{1-\rho}\right) + S_w\left(L_r - \frac{y}{r^2}\right)\right]$$

$$K = \frac{I'}{(1-\rho)}(rL_r + m - wL_w)$$

$$E_\lambda = \frac{\partial(1-\rho)}{\partial(1-\lambda)}\frac{(1-\lambda)}{(1-\rho)}; \quad E_p = \frac{\partial(1-\rho)}{\partial P}\frac{P}{(1-\rho)}$$

Note that $E_\lambda, E_p > 0$ and, substituting (30) into the definition of E_p,

$$E_p = \frac{\lambda B}{Py}\frac{y}{[(1-\gamma c)y + B/P - (1-\lambda)B/P]} = \frac{\lambda B}{(1-\gamma c)Py + \lambda B} < 1$$

---

may be derived as follows. Equilibrium in the presence of an unchanged r under these conditions requires that P change such that $m=(1-\lambda)(1-\rho)B/Pr$ is unchanged. That is, we must have

$$0 = dm = \frac{\partial m}{\partial(1-\lambda)}d(1-\lambda) + \frac{\partial m}{\partial P}dP$$

$$= \frac{B}{r}\left\{\frac{1}{P}\left[(1-\lambda)\frac{\partial(1-\rho)}{\partial(1-\lambda)} + (1-\rho)\right]d(1-\lambda) + (1-\lambda)\left[\frac{P\frac{\partial(1-\rho)}{\partial P} - (1-\rho)}{P^2}\right]dP\right\}$$

or

$$\frac{dP}{d(1-\lambda)} = \frac{P}{(1-\lambda)}\frac{(1+E_\lambda)}{(1-E_p)}$$

where E_λ and E_p are the elasticities of response of $(1-\rho)$ to variations in $(1-\lambda)$ and P and are stated precisely in the Appendix. Equation (A3) reduces to the above expression when $I'=0$.

If we assume P and $r=i(1-\rho)$ to respond positively to the excess demands for commodities and money, respectively, and following the standard analysis of dynamic adjustment processes, it may be shown that a necessary condition for stability is that the denominator of (A1)–(A3) be positive.[19]

It is easily seen from (A1)–(A2) that open-market operations influence both r and i under these conditions even in the absence of wealth effects. That is, $dr/d(1-\lambda)>0$ and $di/d(1-\lambda)<0$ even if $L_w=S_w=0$, where $S_r>0$ and $I',L_r<0$. None of these qualitative results is affected by the adoption of a classical transactions money demand $M/P=ky$, as may be seen by letting $L_r=L_w=0$.

[19]See Metzler (1951, pp. 115–16), Paul Samuelson (pp. 260–65), and Patinkin (1965, pp. 484–94). The denominator of (A1)–(A3) reduces precisely to Patinkin's necessary condition for "the stability of the system as analyzed in the commodity and money markets" if, like Patinkin, we neglect the government budget restraint (i.e., if we let $E_p=\rho=0$) and do not count capitalized income as part of wealth.

REFERENCES

A. S. Blinder and R. M. Solow, "Does Fiscal Policy Matter?," *J. Publ. Econ.*, Nov. 1973, *2*, 319–37.

R. S. Boyer, "Commodity Markets and Bond Markets in a Small, Fixed-Exchange-Rate Economy," *Can. J. Econ.*, Feb. 1975, *8*, 1–23.

______, "Devaluation and Portfolio Balance," *Amer. Econ. Rev.*, Mar. 1977, *67*, 54–63.

W. H. Branson and R. L. Teigen, "Flow and Stock Equilibrium in a Dynamic Metzler Model," *J. Finance*, Dec. 1976, *31*, 1323–40.

R. Clower, "The Keynesian Counterrevolution: A Theoretical Appraisal," in Frank H. Hahn and Frank P. R. Brechling, eds., *The Theory of Interest Rates*, London 1965.

Irving Fisher, *The Theory of Interest*, New York 1930.

John G. Gurley and Edward S. Shaw, *Money in a Theory of Finance*, Washington 1960.

Gottfried Haberler, *Prosperity and Depression*, 3rd ed., Geneva 1941.

______, "The Pigou Effect Once More," *J. Polit. Econ.*, June 1952, *60*, 240–46.

J. R. Hicks, *Value and Capital*, 2d ed., Oxford 1946.

______, *Critical Essays in Monetary Theory*, Oxford 1967.

George Horwich, "Real Assets and the Theory of Interest," *J. Polit. Econ.*, Apr. 1962, *70*, 157–69.

______, *Money, Capital, and Prices*, Homewood 1964.

H. G. Johnson, "Monetary Theory and Policy," *Amer. Econ. Rev.*, June 1962, *52*, 335–84.

E. Karni, "Inflation and Real Interest Rate: A Long-Term Analysis," *J. Polit. Econ.*, Mar./Apr. 1972, *80*, 365–74.

J. M. Keynes, *The General Theory of Employment, Interest and Money*, London 1936.

L. A. Metzler, "Wealth, Saving, and the Rate of Interest," *J. Polit. Econ.*, Apr. 1951, *59*, 93–116.

______, "The Structure of Taxes, Open-Market Operations, and the Rate of Interest," in his *Collected Papers*, Cambridge 1973.

John Stuart Mill, "Of the Influence of Consumption upon Production," in his *Essays on Some Unsettled Questions of Political Economy*, London 1844.

______, *Principles of Political Economy*, London 1848.

Robert A. Mundell, "The Public Debt, Corporate Income Taxes, and the Rate of Interest," *J. Polit. Econ.*, Dec. 1960, *68*, 622–26.

______, "Inflation and Real Interest," *J. Polit. Econ.*, June 1963, *71*, 280–83.

______, *Monetary Theory*, Pacific Palisades 1971.

J. Niehans, "A Comment on Stabilization Paradoxes," in Horst Albach et al., eds., *Quantitative Wirtschaftsforschung*, forthcoming.

______, "Metzler, Wealth, and Macroeconomics: A Review," *J. Econ. Lit.*, Mar. 1978, *16*, 84–95.

Don Patinkin, *Money, Interest and Prices*, 2d ed., New York 1965.

______, "Money and Wealth: A Review

Article," *J. Econ. Lit.*, Dec. 1969, *7*, 1140–160.

Boris P. Pesek and Thomas R. Saving, "Monetary Policy, Taxes, and the Rate of Interest," *J. Polit. Econ.*, Aug. 1963, *71*, 347–62.

______ **and** ______, *Money, Wealth and Economic Theory*, New York 1967.

E. S. Phelps, "Anticipated Inflation and Economic Welfare," *J. Polit. Econ.*, Feb. 1965, *73*, 1–17.

A. C. Pigou, *Employment and Equilibrium*, London 1941.

______, "The Classical Stationary State," *Econ. J.*, Dec. 1943, *53*, 342–51.

Paul A. Samuelson, *Foundations of Economic Analysis*, Cambridge 1947.

T. Scitovsky, "Capital Accumulation, Employment and Price Rigidity," *Rev. Econ. Stud.*, Feb. 1941, *8*, 69–88.

J. L. Stein, "Money and Capacity Growth," *J. Polit. Econ.*, Oct. 1966, *74*, 451–65.

J. E. Stiglitz, "Taxation, Corporate Financial Policy, and the Cost of Capital," *J. Publ. Econ.*, Feb. 1973, *2*, 1–34.

______, "The Corporation Tax," *J. Publ. Econ.*, May 1976, *5*, 303–11.

Henry Thornton, *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, London 1802; reprinted 1939.

R. N. Waud, "Inflation, Unemployment, and Economic Welfare," *Amer. Econ. Rev.*, Sept. 1970, *60*, 631–41.

K. Wicksell, "Taxation in the Monopoly Case," translated by Richard Musgrave and reprinted in Musgrave and Carl S. Shoup, eds., *Readings in the Economics of Taxation*, Homewood 1959.

Inflation Expectations and Money Growth in the United States

By DONALD J. MULLINEAUX*

While inflation expectations have recently come to play a major role in both macroeconomic theorizing and stabilization-policy analysis, relatively little work has been done to date on testing models that purport to explain how forecasters form inflation anticipations. This paper attempts to begin to fill that lacuna by exploiting data on expected inflation recently published by John Carlson (1977a). These data are briefly described in Section I. Section II examines the relationship between inflation expectations and a number of factors commonly believed to influence the actual rate of inflation, such as the money growth rate, a number of fiscal-policy-related variables, and the unemployment rate. Also considered is the question of whether the expectations generating process has "shifted" over time. Section III examines the question of whether forecasters efficiently employ available information in generating inflation predictions. The final section summarizes the results and considers the policy-related implications.

I. Carlson's Version of the Livingston Data

Since 1947, Joseph Livingston has conducted a semiannual price-expectations survey of a relatively small group of business and academic economists. Although a number of researchers have employed the Livingston data as observed measures of inflation expectations (see William Gibson, Kajal Lahiri, Timothy McGuire, D.H. Pyle, Stephen Turnovsky, and Turnovsky and Michael Wachter), the notion that these data are "suspect" appears to be quite common among economists. Indeed, James Pesando explicitly rejects the Livingston data as valid proxies for "the market's" expectations on the grounds that the data fail a test for rationality in the sense of John Muth. Carlson (1977a) presents a more telling criticism of the Livingston data. He notes that the data employed by most researchers has in some instances been judgmentally adjusted by Livingston in an attempt to take account of information unavailable when the forecast was made.[1] Carlson notes that these adjustments do not appear to have been applied consistently and hence the numbers published by Livingston may well be contaminated by measurement errors. Carlson employs the unadjusted mean values of the survey responses to create a consistent series that takes account of the information problems that led Livingston to adjust the data. I have recently demonstrated that over the period 1959–69, Carlson's expected inflation series does exhibit a weak form of Muthian rationality.

*Vice president and associate director of research, Federal Reserve Bank of Philadelphia. I am grateful to Herb Taylor who provided first-rate research assistance, and to James O'Brien, Robert Shiller, an anonymous referee, and the managing editor for helpful comments. The usual disclaimer applies.

[1]Each December, Livingston publishes expected *levels* of the Consumer Price Index (*CPI*), and Wholesale Price Index for June and December of the following year. In June, he presents forecasts for December and the following June. Survey results are usually collected the month before publication and the most recent observation available when the forecast is made is that *two* months before publication (see Carlson, 1977a). Whenever the price index changes between the survey date and publication date, Livingston *sometimes* adjusts the mean forecast to reflect the new information. Most researchers uncritically employ the adjusted forecast levels to calculate six-month and twelve-month inflation rates. This procedure is valid only if Livingston's adjustments accurately and consistently capture the effect of new information on inflation forecasts. Carlson (1977a) argues the more appropriate procedure is to use unadjusted means and calculate eight-month and fourteen-month inflation forecasts.

In what follows I employ Carlson's series of expected *CPI* inflation over a six-month horizon[2] to test a number of hypotheses concerning the formation of inflation expectations. The data are biannual observations with a forecast base dated either in April or October of each year. Though the Carlson series ends in 1975, Livingston was kind enough to make the raw data available for the 1976 and 1977 surveys. Carlson's procedure was employed to generate expected *CPI* inflation rates, so that our sample contains four additional observations relative to that published in Carlson (1977a).

II. Money Growth in an Inflation Expectations Model

A. *Problems with Previous Research*

The "standard" approach to modelling inflation expectations is to utilize a distributed lag in past rates of inflation (see Jacob Mincer and Stanley Diller). Forecasters are thus presumed to consider a relatively narrow information set when generating inflation predictions. Such models are frequently criticized (see John Rutledge and McGuire for examples) on the grounds that they ignore variables that economic theory suggests are relevant to the determination of the inflation rate. However, empirical tests of the proposition that inflation expectations are affected by other factors, such as the rate of money growth, suffer several shortcomings. The most important of these is that researchers have failed to utilize "good" measures of inflation expectations.[3] In the next section, I use Carlson's version of the Livingston data to attempt to determine whether forecasters do consider money growth information in generating forecasts of inflation.

B. *Do Forecasters Process Money Growth Information?*

It is now well known that extrapolation of the past history of a forecast variable yields efficient (rational) predictions only in a special case. In particular, if the stochastic process generating a random variable can be expressed as a weighted sum of a univariate random shock, then the minimum-mean squared error (*MMSE*) forecast can be obtained via extrapolation (see Nelson). Where imperfectly correlated multiple shocks generate a random variable, however, forecasters can increase efficiency (reduce *MMSEs*) by considering the past history of *all* the relevant shocks. Of course, this need not imply that it is "economically rational" to process all relevant information in forming a prediction. Information gathering and processing is costly, and hence the optimal information set will depend on the marginal benefit of more information (in reduced-mean square forecast errors) relative to its marginal cost. Edgar Feige and Douglas Pearce have suggested that forecasters should not consider information on past money growth rates because they cannot reduce the variance of their prediction

[2]Carlson (1977a) actually calculates an expected inflation rate over an eight-month horizon. If we assume, however, that *steady* inflation is expected over both the six- and eight-month horizons, the forecasts when expressed as annual rates will be the same over both time periods. We utilize the shorter time horizon to avoid any overlap in the time periods of successive forecasts. This allows us to make the convenient assumption (for estimation purposes) that the time dimension for the "information set" variables corresponds to that of the dependent variable, for example, that they are six-month rates of changes running from April to October and from October to April.

[3]McGuire, for example, employed the "old" version of the Livingston series, which, as noted above, is suspect on several counts. Rutledge, on the other hand, used an estimate of the forward rate of interest on three-month Treasury Bills as a proxy for inflation expectations. To interpret the forward rate this way requires that: 1) the real rate of interest is constant, and 2) any liquidity premium embedded in forward rates is also constant over time. Though Eugene Fama (1975) has presented evidence to support the constant real rate proposition, his results have been challenged by Charles Nelson and G. William Schwert, Carlson (1977b), and Kenneth Garbade and Paul Wachtel. In addition, Fama (1976) has generated separate evidence that liquidity premia are *not* constant over time. Thus, neither of Rutledge's maintained hypotheses is supported by strong evidence. While movements in forward rates may well be related to changes in money growth, there is little reason to conclude that this solely reflects a relation between money growth and inflation expectations.

errors by looking at these data. Their results are controversial, however. I address the same issue by asking whether forecasters' behavior suggests they consider money growth information. The issue of whether money growth "should be" an information variable is considered in Section III. Testing that hypothesis requires that we address the somewhat complicated issue of describing money growth "information."

Presumably forecasters would want to consider "up-to-date" information at the time of their predictions. Some judgements are therefore required about what data were available at the time predictions were generated. I employed two different assumptions concerning "available" information on money growth. In one instance, it was assumed that inflation forecasters were aware of the most recent six-month growth rate that could be calculated from money supply data published in the *Federal Reserve Bulletin* for the months in which the forecasts were generated.[4] I call this the "more up-to-date" data series and note that it involves several shifts in the dates of the six-month horizon over which money growth is calculated, since the Fed has varied the speed with which money supply data is published over time. I also constructed a "less up-to-date" data series which fixes the horizon over which money growth is calculated, so that the end month is always that two months prior to the month in which the forecast is made. I calculated both of these series for both seasonally adjusted and unadjusted data. Seasonally adjusted data have been published by the Fed since 1955, while unadjusted data have been available for a longer period. Since there appear to be a number of problems associated with the Fed's seasonal adjustment procedures, forecasters might believe there is more useful information in the unadjusted data. This issue, as well as the question of the time horizon for money growth information, can be investigated empirically by estimating separate equations using different data and comparing the properties of the different equation estimates.

There is still another data problem to be addressed, however. Forecasters might also be interested in the money growth rates prior to the six-month period immediately preceding the forecast. To test this hypothesis, one could simply lag the money growth rates calculated as described above. However, if revisions in the data had occurred and if forecasters were aware of them, this procedure introduces measurement error. This would bias the coefficients of lagged money growth rates toward zero, if errors in money growth are independent of errors in measuring inflation expectations.

Investigating these issues involved experimenting with a large number of regressions. Rather than report all the findings, I will simply note that the results are rather insensitive to the issue of the timing of the money growth information. The adjusted R^2 never differed by more than a percentage point when the two different money growth information time horizons were compared. Nor could I uncover any significant differences in the results when adjustments were made to lagged money growth information to take account of revisions in the data. In what follows, I present results using the more up-to-date money growth information set and past money growth will consist of lagged values of this series.

I report ordinary least squares (*OLS*) results of estimates of the following two equations:

$$(1) \quad p_t^e = a + \sum_{i=1}^{2} b_i p_{t-i} + \sum_{j=1}^{2} c_j m_{t-j} + u_t$$

$$(2) \quad p_t^e = a + \sum_{k=1}^{5} r_k p_{t-k} + q m_{t-1} + e_t$$

where p_t^e is the current period's forecast of *CPI* inflation over the next six months, p_{t-1} is the most recently observed six-month inflation rate, m_{t-1} is the most recently observed six-month money growth rate (seasonally adjusted), and the additional variables are lagged values of the relevant

[4] In generating his inflation forecast series, Carlson presumed that forecasts were generated every May and November with *CPI* price levels in April and October representing the most recent information. Whether forecasters possessed information on the level of the money supply for these same months is unclear.

TABLE 1—*OLS* ESTIMATES OF EQUATIONS (1) AND (2), WITH SEASONALLY ADJUSTED MONEY GROWTH DATA: 1956–77

Variable	Equation (1)[a]	Equation (2)[a]	Equation (1)	Equation (2)
Intercept	0.045	−0.202	−0.541[b]	−0.505[b]
	(0.20)	(−1.10)	(−2.83)	(−3.20)
p_{t-1}	0.391[b]	0.402[b]	0.287[b]	0.349[b]
	(5.25)	(6.70)	(4.96)	(7.21)
p_{t-2}	0.292[b]	0.169[b]	0.286[b]	0.216[b]
	(3.93)	(2.35)	(5.29)	(3.77)
p_{t-3}	–	−0.040	–	−0.085
	–	(−0.54)	–	(−1.43)
p_{t-4}	–	0.025	–	0.042
	–	(6.35)	–	(0.73)
p_{t-5}	–	0.223[b]	–	0.167[b]
	–	(3.72)	–	(3.44)
m_{t-1}	–	–	0.186[b]	0.165[b]
	–	–	(4.85)	(4.92)
m_{t-2}	–	–	0.090[b]	–
	–	–	(2.13)	–
$\bar{R}^2$	.819	.889	.904	.931
S.E.	0.874	0.685	0.636	0.540
D.W.	0.650	1.017	0.867	1.184
h	5.280	3.436	4.037	2.559
F^*	–	–	19.213	24.216
Critical *F* (.05)	–	–	3.24	4.11

[a]The coefficients of the money growth terms are constrained to zero.
[b]Indicates the coefficient is significantly different from zero at the .05 level.

terms. All variables are expressed as annual rates. The terms u_t and e_t are random error terms assumed to possess the well-known classical properties. The M_1 definition of money is employed; that is, money is defined as that sum of currency and demand deposits. The same equations were also estimated with unadjusted money growth replacing seasonally adjusted money growth. However, none of the experiments carried out in this paper disclosed substantively different results when unadjusted data were employed. Hence, to conserve space, only the results based on seasonally adjusted data are reported here. The results are shown in Table 1, along with estimates of the same equations with the money growth terms deleted. The figures in parentheses in the table are *t*-ratios, *S.E.* is the standard error of estimate, *D.W.* is the Dubin-Watson statistic, and *h* is the statistic suggested by Durbin for testing the hypothesis of first-order serial correlation in the presence of a lagged dependent variable. The F^*-values are calculated in a test of the hypothesis that all the elements of the set of money growth coefficients equal zero.

I experimented with a number of other specifications involving longer lags in past inflation and/or past money growth, but the additional terms were never significant by the usual *t*-tests or an *F*-test. Inspection of Table 1 shows that inflation forecasts are explained a little better by two lagged terms each in past inflation and past money growth than by a five-term lag in past inflation alone. I present several specifications, since, as will become clear in the next part of this paper, the structure of the inflation-forecast-generating process does not appear to have been stable over time. Identifying superior models is therefore considerably complicated.

For our purposes, the most important result observed in this stage of the estimation process is that money growth information contributes significantly to the explanation of inflation expectations, given the effect of past inflation on anticipations. The F^*-values for testing the hypothesis that all the money growth coefficients simultaneously equal zero can be rejected in every case.

There is a serious shortcoming in these results, however. Even with the inclusion of

several variables that are close analogues to lagged dependent variables, the $D.W.$ statistics are quite low. Indeed, the values of the h-statistic suggest we cannot reject the hypothesis of first-order serial correlation in the errors of *any* of the equations. This observed behavior of the residuals may be related to shifts in the structure of the model over time. Indeed, Christopher Sims argues that since expectations are at least partly rational, "we must accept the idea that what is called the structure in textbook treatments of simultaneous equation models can change under policy changes which affect only the time paths of the exogenous variables" (p. 294). Robert Lucas also has stressed the point that the structure of expectations formation equations should be functionally related to perceived policy changes.

C. *Has the Inflation Expectations Equation Shifted Over Time?*

I conducted a rough test for structural change by arbitrarily breaking all the samples into two periods using 1966 as the division point. Using the Chow test, I rejected the hypothesis of model homogeneity at the .05 level for all the equations containing money growth variables.[5] The time-disaggregated equations also showed lower values of the h-statistic, suggesting that structural change could indeed be a factor accounting for the high h in the aggregated sample. (See Table 2 for the Chow-test F-values and the h-statistics for the disaggregated equations.) This procedure is clearly unsatisfactory for estimating the impact of money growth on expectations, however.

[5]The equations containing *only* past inflation terms also disclosed structural shifts over the same time periods. When tested over the periods 1949–65 and 1966–77, the Chow F-value for an equation with five lagged inflation terms was 10.84 (critical $F_{.05}=2.30$), while the equation with two lagged inflation terms showed a Chow-test value of 16.86 (critical $F_{.05}=2.79$). Similarly large F-values were observed for the periods 1956–65 and 1965–77. Other researchers who have employed the Livingston data, such as Gibson, Turnovsky, and Lahiri, have reported findings of structural shifts in inflation expectations equations.

TABLE 2—CHOW TEST VALUES AND h-STATISTICS FOR DISAGGREGATED EQUATIONS

	F-Values		h-Values	
Equation Specification	Chow-Test Statistic	Critical F (.05)	1956–65 Period	1966–77 Period
$2p_t$ terms, $2m_t$ terms	2.58	2.49	0.68	2.67
$5p_t$ terms, $1m_t$ term	5.20	2.34	0.11	0.77

There is no reason to believe that a discrete shift occurred only in 1966, and that the model coefficients are constant over the two subperiods. In fact, some of the h-statistic values remain high enough in the disaggregated equations that we still cannot reject the autocorrelation hypothesis. A more general hypothesis is that the coefficients vary in every period. Such a model can be tested using the varying-parameter regression (VPR) technique developed by Thomas Cooley and Edward Prescott.

To estimate a relation with time-varying parameters, assume that the coefficients (or a subset thereof) are subject to both permanent and transitory changes over time:

$$B_t = B_t^p + u_t \tag{3}$$

$$B_t^p = B_{t-1}^p + w_t \tag{4}$$

The vector of B_t^p represents the permanent component of the parameters at time t. The u_t and w_t are independent and identically distributed random variables with zero mean vectors and covariance matrices as follows:

$$cov\ (u_t) = (1-\gamma)\sigma^2\Sigma_u \tag{5}$$

$$cov\ (w_t) = \gamma\sigma^2\Sigma_w \tag{6}$$

The matrices Σ_u and Σ_w specify the relative magnitudes of the variances of the parameter changes and are assumed known up to a scale factor.

The parameter γ (which is constrained to lie between zero and one) specifies the relative variance of the permanent and transitory components of the change in the Bs. The greater the value of γ, the more important is permanent relative to transitory

change. The object of the estimation procedure is to obtain a consistent value of γ which will yield the asymptotically efficient estimates of the Bs.

The process generating the parameters is nonstationary; hence it is impossible to specify a likelihood function. However, the likelihood function conditional on a specific realization of the parameter process is well-defined. Using the conditional likelihood function amounts to treating the unknown parameters as random, but stationary at a particular point in time. If we let T denote such a period and X_t the row vector of information variables, we can write the inflation expectation equation as

$$p_t^e = X_t' B_T + \pi_t \tag{7}$$

The error vector π is distributed normally with mean zero and covariance matrix:

$$cov(\pi) = \sigma^2\left[(1-\gamma)R + \gamma Q\right] \equiv \sigma^2 \Omega(\gamma) \tag{8}$$

The matrix R is a diagonal matrix which depends on Σ_u and X, while Q is a matrix which depends on X, Σ_w, and the period on which the parameter process is normalized.

In what follows, I calculated Bayesian estimates of the parameters, assuming that prior knowledge about the parameters B, σ, and γ can be represented by the following locally uniform, independent distributions:

$$\begin{aligned} p(\gamma) &= d\gamma \qquad 0 \leqslant \gamma \leqslant 1 \\ p(B) &\propto k_1 \\ p(\sigma) &\propto 1/\sigma \, d\sigma \end{aligned} \tag{9}$$

Let B_γ be the Aitken estimator of B:

$$B_\gamma = \left(X' Q_\gamma^{-1} X\right)^{-1} X' Q_\gamma^{-1} p^e \tag{10}$$

and S_γ be the generalized sum of squared residuals:

$$S_\gamma = (p^e - XB_\gamma)' Q_\gamma^{-1} (p^e - XB_\gamma) \tag{11}$$

Following Cooley and Prescott, the parameters B have the posterior density

$$p(B; p^e, \gamma, \sigma) \sim N\left[B_\gamma, \sigma^2\left(XQ_\gamma^{-1}X\right)^{-1}\right] \tag{12}$$

and the marginal posterior for γ is

$$p(\gamma, p^e) \propto |Q_\gamma|^{-\frac{1}{2}} \left|\left(XQ_\gamma^{-1}X\right)^{-1}\right|^{\frac{1}{2}} S^{-(T-k)/2} \tag{13}$$

Conditional on γ then, the first moment of B is simply the Aitken estimator B_γ, which can be found via numerical integration.

An advantage of the Bayesian estimation procedure is that it allows us to utilize the calculation of posterior odds (see Arnold Zellner, pp. 291–317). In general, the posterior odds depend on the prior odds ratio, the nature of the assumed prior distributions, and the sample information regarding the goodness of fit of the competing models. However, since we have assumed equal prior odds and diffuse priors that do not vary from one experiment to another, the posterior odds depend strictly on the relative likelihood ratios (see Zellner, pp. 310–12). Hence, choosing the model with higher posterior probability is compatible with the rule of choosing the model with the higher R^2. Model selection is, of course, quite complicated here, given the large variety of patterns of parameter variation that can be assumed. I do not claim to have identified the "true" or "best" model of inflation expectations, but neither do I believe the results are totally arbitrary since the data have been allowed to disclose which specification is most likely to have generated the observations in question.

The model which yielded the highest posterior odds was one which involved permanent and transitory change in all the coefficients, including the intercept.[6] My specification of the covariance matrices was $\Sigma_u = \Sigma_w = \sigma^2 I$.

[6] I also considered models in which parameter change was restricted to the money growth coefficients, then only to the past inflation coefficients, and only to the intercept. Assuming prior odds of .50 –.50, none was superior on the basis of calculation of posterior odds to the model in which all the coefficients varied. The posterior odds on the model we report relative to an *OLS*-estimated constant coefficient model exceed .99999–.00001 (assuming prior odds of .50 –.50). It should also be noted that the results were relatively insensitive to shifts on the assumed values on the diagonal of the Σ_u and Σ_w matrices (as long as the assumed values were nonzero).

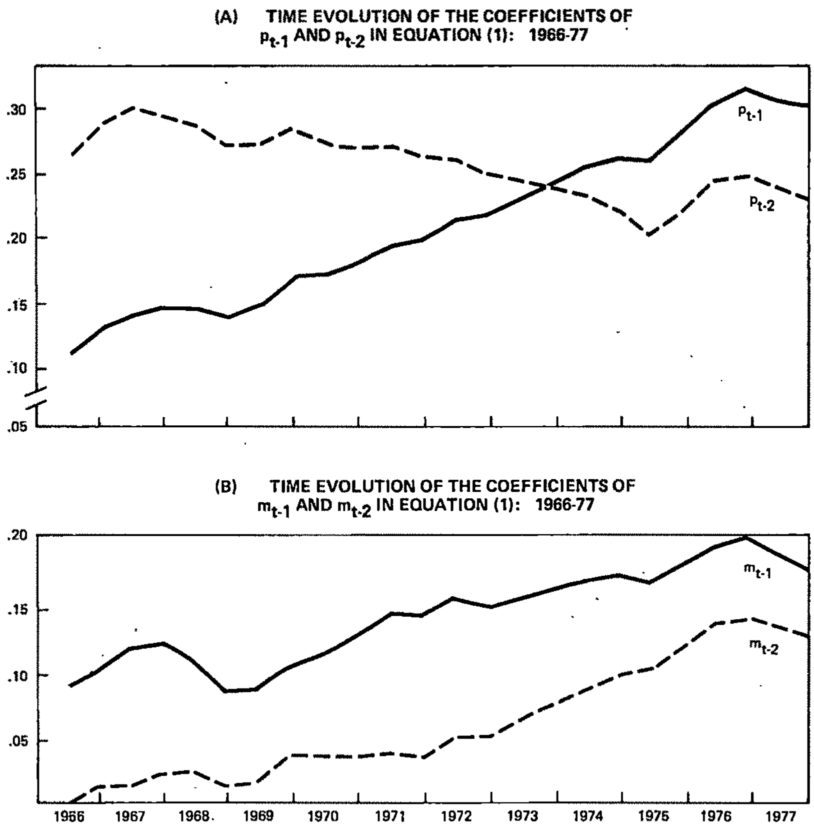


Figure 1

The justification for allowing the intercept to vary is that our sample information represents a time-series of averages of cross-section survey responses in which the respondents varied over time. Hence, the error term in our model captures both individual-specific and period-specific effects. Permitting intercept variation appears to be an efficient means of estimating this kind of "variance components" model.

For each of the specifications discussed above, I estimated the permanent values of the coefficients in every period starting from 1966 and ending in 1977. The time evolution of the estimated permanent values of the coefficients in Equation (1) are shown in Figure 1. Figure 2 shows the evolution of the money growth coefficient and the coefficient of the initial term in the lag in past inflation rates from equation (2); the other coefficients varied relatively little and are not shown here. The estimate of γ was .62 for equation (1) and .61 for equation (2), indicating that just over 60 percent of the observed variation in the coefficient values was estimated to be permanent in each specification.[7] The results of the varying-parameter-estimation process are thus consistent with the Lucas-Sims proposition that the structure of a model of inflation expectations is functionally related to perceived policy changes.

[7]The estimates of the standard error of γ were 0.221 for equation (1) and 0.211 for equation (2). The ratio of the estimated γ to its estimated standard error is asymptotically normal (see Cooley and Prescott); employing the asymptotic distribution permits rejection of the hypothesis that $\gamma=0$ at the .01 level.

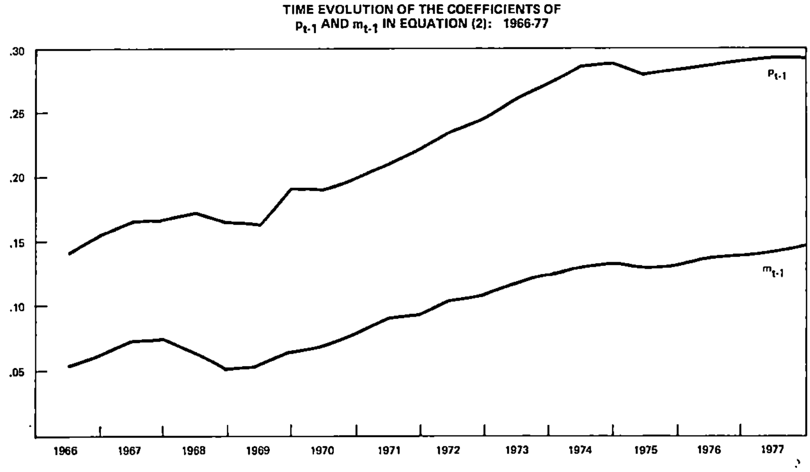


FIGURE 2

D. *Do Other Variables Influence Inflation Expectations?*

The model on which these results are based is restrictive in assuming that the *only* information that forecasters consider is observed inflation rates and money growth rates. Inflation forecasts may well depend on other factors, such as the observed pattern of government spending or the federal deficit, or on other variables believed to influence aggregate demand. Forecasters might perceive the existence of a *direct* relation between, say, the current size of the federal deficit and future inflation. Alternatively, they could believe the deficit is linked indirectly to inflation via its impact on future money growth. I did not attempt to distinguish among such influences empirically, however. Rather I chose to investigate whether the data were consistent with the existence of *either* type of relationship.

The "information variables" examined included the unemployment rate, growth in federal government spending on goods and services, the federal deficit, and the level of federal government spending as a percentage of *GNP*. Each of these variables was entered in distributed lag form into equations (1) and (2). Different lag lengths were examined, including cases wherein only the most recently observed value of the variable was included in the regression. I investigated *OLS* estimates of regressions over the full sample horizon as well as over several subsample periods, and in no case did I uncover a significant relationship between any fiscal-related variables and inflation expectations. The same result obtained when a number of similar specifications were estimated with the varying-parameter regression (*VPR*) technique. Money growth continued to have a significant impact on inflation expectations when government spending variables were included in the regressions, however, and the coefficient of the money growth variables showed a pattern of evolution similar to that reported above.

The impact of information about unemployment on inflation expectations was less straightforward. When two unemployment terms[8] were included in equations (1) and (2), the coefficient of each was significant, but a somewhat curious pattern of signs was observed (see Table 3). An increase in the most recently observed unemployment rate has a negative effect on inflation forecasts, whereas a rise in unemployment in the prior

[8]In experiments with longer lags in unemployment, the additional terms were not significant by the *F*-test criterion. In fact, when longer lags were estimated, *all* the coefficients of the distributed lag failed the test for significance.

TABLE 3—*OLS* ESTIMATES OF EQUATIONS (1) AND (2) WITH UNEMPLOYMENT TERMS INCLUDED (*unp*)

Variable	Equation (1) 1956–77	Equation (2) 1956–77	Equation (1) 1956–72	Equation (2) 1956–72
Intercept	-1.745^a	-1.045^a	-0.815	1.221
	(−4.90)	(−2.38)	(−1.19)	(1.30)
p_{t-1}	0.326^a	0.311^a	0.246^a	0.128
	(6.75)	(6.92)	(3.37)	(0.089)
p_{t-2}	0.300^a	0.271^a	0.286^a	0.158
	(5.70)	(4.90)	(3.99)	(1.87)
p_{t-3}	–	−0.038	–	−0.029
	–	(−0.67)	–	(−0.42)
p_{t-4}	–	0.042	–	0.136
	–	(0.78)	–	(1.81)
p_{t-5}	–	0.094	–	0.168^a
	–	(1.82)	–	(2.49)
m_{t-1}	0.155^a	0.149^a	0.157^a	0.133^a
	(4.87)	(4.79)	(4.36)	(3.66)
m_{t-2}	0.041	–	0.071	–
	(1.02)	–	(1.58)	–
unp_{t-1}	−0.268	-0.345^a	−0.145	−0.371
	(−1.65)	(−2.32)	(−0.74)	(−2.13)
unp_{t-2}	0.515^a	0.450^a	0.234	0.110
	(3.30)	(3.04)	(1.05)	(0.52)
$\bar{R}^2$	.937	.943	.813	.825
S.E.	0.513	0.493	0.532	0.494
D.W.	1.227	1.208	1.081	0.907
h	2.544	2.477	3.041	3.788
F^*	11.439	4.683	0.605	3.035
Critical *F* (.05)	3.25	3.27	3.33	3.37

[a] Indicates the coefficient is significantly different from zero at the .05 level.

six-month period *increases* anticipated inflation.[9] This result, however, is not robust with respect to the selection of the sample period. If the 1973–77 period is excluded from the sample, we cannot reject the hypothesis that unemployment information makes no significant contribution to explaining inflation expectations (see Table 3).[10] Money growth information remains significant, however, in each case. I also estimated the same equations over the full sample period using *VPR* techniques. A comparison of posterior odds disclosed that the model with unemployment as a variable was distinctly inferior to the model which contained only past inflation and money growth information. In addition, the pattern of coefficient variation reported above was virtually unaltered when unemployment was entered as a variable. My conclusion is that the hypothesis that inflation expectations are systematically related to unemployment information gets only weak support from the data, but that the question deserves additional research. In particular, the existence of an *indirect* link between unemployment and inflation expectations via shifts in *anticipated* stabilization policy actions should be investigated.

III. Are Inflation Forecasts Rational?

The justification for including money supply growth information in an equation explaining inflation expectations is, of course, the notion that *actual* inflation is influenced by the recent pattern of money growth. In this section, we will test that

[9]One way to rationalize such a result would be to posit that forecasters envision a lagged response by the Federal Reserve to an observed increase in unemployment in the form of an increase in the money growth rate; i.e., that the lagged unemployment rate is a proxy for *anticipated* rather than observed money growth.

[10]The same conclusion holds if the 1974–77 period is the one excluded from the sample.

proposition. In a sense, we are asking not whether money growth *is* an information variable, but whether it *should be* one. We need to investigate this issue because we want to examine whether forecasters efficiently incorporate money growth information into their predictions. In particular, we want to know if the inflation *expectations* model is identical to the process determining the *actual* inflation rate (apart from a random error), that is, if the Livingston forecasts are rational in the sense of Muth. Judging the results of such a test requires that we have some knowledge of whether the actual inflation rate is related to observed money growth rates.

Table 4 presents the results of the *OLS* estimation of the same set of equations presented in Table 1 except that the actual inflation rate over the forecast horizon replaces the Livingston data prediction as the dependent variable. The results indicate that in one of the two cases (equation (2)), we can reject the hypothesis that money growth fails to contribute toward explaining the actual inflation rate at the .05 level. When a shorter lag on past inflation is employed, however, the hypothesis that "money growth doesn't matter" cannot be rejected at the .05 level, but it can be at the .10 level.[11] Though the evidence is less than overwhelming, I conclude that it is sufficiently strong to justify a straightforward test of the rationality proposition.

To conduct such a test, I regressed the inflation forecast *error* on the elements of the information set used to forecast. Efficient forecasting requires that the coefficients of all the information variables equal zero. A nonzero coefficient implies that forecasters could have improved their predictions by better exploiting the information set (or a subset thereof). Rationality also requires that the error term in this equation be serially uncorrelated. Autocorrelated forecast errors would imply that predictions could be improved by simply taking account of this phenomenon in generating predictions.

The results of regressions in which inflation forecast errors are regressed on the information sets assumed in equations (1) and (2) are presented in Table 5. Both of the $\bar{R}^2$ are quite low, and in each case we cannot reject the hypothesis that the coefficients of all the information-set variables equal zero (see Table 5 for *F*-values). The value of the Durbin *h*-statistic suggests the absence of first-order serial correlation in the residuals, though in one case *h* could not be calculated. Hence, the evidence does not enable us to reject the rationality hypothesis.[12]

Estimating these relations by *OLS* involves assuming a constant structural relation for the forecast error. This implies that the *actual* inflation process underwent structural changes that were not systematically different from what we observed for the expectations-generating function. I tested the hypothesis that the structure of the forecast-error equation was stable over time employing both the Chow test and *VPR* methods. I could not reject the hypothesis for any of the specifications employed.

IV. Summary and Conclusions

This paper has addressed the following questions: 1) what information do forecasters employ in generating inflation predictions; 2) is the relationship between inflation expectations and information variables temporally stable; 3) are the forecasts rational in the sense that they efficiently incorporate "available" information?

[11]These equations were estimated using the "as initially published" data on money growth employed above. If the same equations are estimated with *revised* data (as of June 1978) on money growth, then the hypothesis that the money growth coefficients all equal zero is rejected in every case. For seasonally adjusted data, for example, the *F*-values are 4.404 for equation (1) and 10.124 for equation (2). While the usefulness of money growth information for forecasting clearly should be investigated with as initially published data, most research into the relationship between inflation and money growth employs revised data. To the extent that forecasters rely on such results to identify information variables, they will overestimate the usefulness of information on money growth for inflation forecasting purposes.

[12]The same results were obtained when the unemployment rate was included as a variable in the information set on which the forecast error was regressed.

TABLE 4—*OLS* ESTIMATES OF EQUATIONS (1) AND (2) USING ACTUAL INFLATION AS THE DEPENDENT VARIABLE
(Seasonally Adjusted Money Growth Data: 1956–77)

Variable	Equation (1)[a]	Equation (2)[a]	Equation (1)	Equation (2)
Intercept	0.912	0.876	0.301	0.379
	(1.94)	(1.82)	(0.57)	(0.76)
p_{t-1}	0.653[b]	0.670[b]	0.534[b]	0.577[b]
	(4.20)	(4.26)	(3.33)	(3.77)
p_{t-2}	0.146	0.261	–	0.364
	(0.93)	(1.36)	–	(1.96)
p_{t-3}	–	−0.092	–	−01.164
	–	(−0.47)	–	(−0.88)
p_{t-4}	–	−0.298	–	−0.308
	–	(−1.51)	–	(−1.66)
p_{t-5}	–	0.277	–	0.220
	–	(1.69)	–	(1.41)
m_{t-1}	–	–	0.199	0.262[b]
	–	–	(1.81)	(2.39)
m_{t-2}	–	–	0.084	–
	–	–	(0.72)	–
$\bar{R}^2$	.597	.603	.626	.648
S.E.	1.804	1.792	1.740	1.687
D.W.	1.755	1.808	1.723	1.751
h	c	c	c	c
F^*	–	–	2.521	5.731
Critical *F* (.05)	–	–	3.23	4.11

[a] See Table 1.
[b] See Table 1.
[c] Cannot be calculated.

TABLE 5–*OLS* ESTIMATES OF REGRESSIONS OF INFLATION FORECAST ERRORS ON THE INFORMATION VARIABLES IN EQUATIONS (1) AND (2)
(Seasonally Adjusted Money Data)

Variable	Equation (1)	Equation (2)
Intercept	0.815	0.859
	(1.55)	(1.84)
p_{t-1}	0.238	0.221
	(1.50)	(1.54)
p_{t-2}	−0.117	0.163
	(−0.78)	(0.94)
p_{t-3}	–	−0.082
	–	(−0.47)
p_{t-4}	–	−0.366
	–	(−2.10)
p_{t-5}	–	0.065
	–	(0.45)
m_{t-1}	0.022	0.107
	(0.21)	(1.04)
m_{t-2}	−0.006	–
	(−0.05)	–
$\bar{R}^2$	−.013	.142
S.E.	1.817	1.580
D.W.	1.647	1.760
h	a	0.581
F^*	0.864	2.170
Critical *F* (.05)	2.62	2.36

[a] Cannot be calculated.

The data indicated that inflation forecasts are systematically influenced by past inflation rates and past rates of money growth, but not by fiscal-policy-related variables such as growth in government spending or the size of the federal deficit. The evidence concerning the relationship between inflation expectations and observed unemployment rates was ambiguous. Unemployment significantly affects the inflation outlook in *OLS* regressions only when the 1973–77 period is included in the sample. In the experiments employing *VPR* estimation techniques, unemployment was never a significant factor. Since the link between unemployment and inflation expectations may be an indirect one, this issue appears deserving of additional research.

The results of traditional covariance analysis as well as *VPR* experiments indicated that the relation between anticipated inflation and the assumed set of information variables has shifted over time. The influence of both money growth and of the most recently observed inflation rate appears to have become systematically more important over time in affecting inflation expectations.

Finally, the evidence supports the view that forecasters efficiently process information on past inflation and money growth in making inflation predictions. We could not reject the hypothesis that Carlson's version of the Livingston data satisfy the property of rationality in the sense of Muth.

My results have some implications for the debate concerning "rational expectations" and the efficacy of monetary policy (see Thomas Sargent and Neil Wallace). The evidence is consistent with the propositions that 1) rational forecasts will take account of policy variables in forming expectations of relevant variables such as the rate of inflation, and 2) if forecasts are based on an unbiased estimate of these variables, then inflation prediction errors will not be systematically biased. To the extent, then, that 1) policy linkages depend on systematic forecast errors, and 2) the forecast errors calculated from the Livingston survey data are valid proxies for the errors made by "market participants," there is no scope for successful countercyclical monetary policy. Research directed toward comparing the properties of the Livingston data with those of other measures of inflation expectations seems desirable as a step toward resolving the latter issue.

REFERENCES

J. A. Carlson, (1977a) "A Study of Price Forecasts," *Annals Econ. Soc. Measure.*, Winter 1977, *6*, 27–56.

______, (1977b) "Short-Term Interest Rates as Predictors of Inflation: Comment," *Amer. Econ. Rev.*, June 1977, *67*, 469–75.

T. Cooley and E. Prescott, "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, Jan. 1976, *44*, 167–84.

S. Diller, "Expectations in the Term Structure of Interest Rates," in Jacob Mincer, ed., *Economic Forecasts and Expectations*, New York 1969.

J. Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, May 1970, *38*, 410–21.

E. F. Fama, "Inflation Uncertainty and Expected Returns on Treasury Bills," *J. Polit. Econ.*, June 1976, *84*, 427–48.

______, "Short-Term Interest Rates as Predictors of Inflation," *Amer. Econ. Rev.*, June 1975, *65*, 269–82.

E. L. Feige and D. K. Pearce, "Economically Rational Expectations: Are Innovations in the Rate of Inflation Independent of Innovations in Measures of Monetary and Fiscal Policy?," *J. Polit. Econ.*, June 1976, *84*, 499–522.

K. Garbade and P. Wachtel, "The Time Variation in the Relationship Between Inflation and Interest Rates," Salomon Brothers Center Stud. Finan. Inst., New York Univ. 1976.

W. E. Gibson, "Interest Rates and Inflationary Expectations," *Amer. Econ. Rev.*, Dec. 1972, *62*, 854–65.

K. Lahiri, "Inflationary Expectations: Their Formation and Interest Rate Effects," *Amer. Econ. Rev.*, Mar. 1976, *66*, 124–31.

Joseph A. Livingston, Surveys published twice yearly, *Philadelphia Sunday Bulletin,* 1948–71; *Philadelphia Inquirer,* 1972 on.

R. E. Lucas, Jr., "Econometric Policy Evaluation: A Critique," in Karl Brunner and Alan Meltzer, eds., *The Phillips Curve and Labor Markets,* Vol. 1, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 19–46.

T. McGuire, "Price Expectations and the Phillips Curve," in Karl Brunner and Alan Meltzer, eds., *The Economics of Wage and Price Controls,* Vol. 2, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 65–114.

J. Mincer, "Models of Adaptive Forecasting," in his *Economic Forecasts and Expectations,* New York 1969.

D. J. Mullineaux, "On Testing for Rationality: Another Look at the Livingston Price Expectations Data," *J. Polit. Econ.*, Apr. 1978, *86*, 329–36.

J. F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica,* July 1961, *29*, 313–35.

C. R. Nelson, "Rational Expectations and Predictive Efficiency of Econometric Models," *J. Bus., Univ. Chicago,* July 1975, *48*, 331–43.

______ **and G. W. Schwert,** "On Testing the Hypothesis that the Real Rate of Interest is Constant," *Amer. Econ. Rev.*, June 1977, *67*, 478–86.

J. Pesando, "A Note on the Rationality of the Livingston Price Expectations Data," *J. Polit. Econ.*, Aug. 1975, *83*, 849–58.

D. H. Pyle, "Observed Price Expectations and Interest Rates," *Rev. Econ. Statist.*, Aug. 1972, *54*, 275–80.

John Rutledge, "*A Monetarist Model of Inflationary Expectations,*" Lexington 1974.

T. Sargent and N. Wallace, "Rational Expectations, the Optimal Monetary Instrument, and The Optimal Money Supply Rule," *J. Polit. Econ.*, Apr. 1975, *83*, 241–54.

C. A. Sims, "Distributed Lags," in Michael Intriligator and David Kendrick, eds., *Frontiers of Quantitative Economics,* Amsterdam 1974.

S. J. Turnovsky, "Empirical Evidence on the Formation of Price Expectations," *J. Amer. Statist. Assn.*, Dec. 1975, *65*, 1441–54.

______ **and M. L. Wachter,** "A Test of the Expectations Hypothesis Using Directly Observed Wage and Price Expectations," *Rev. Econ. Statist.*, Feb. 1972, *54*, 47–54.

Arnold Zellner *An Introduction to Bayesian Inference in Econometrics,*, New York 1971.

Measuring Technological Bias

By RODNEY STEVENSON*

Since J. R. Hicks' *Theory of Wages*, several notions of neutral technological change have appeared in the literature—as well as several interpretations of Hicksian neutrality. Since Hicks, several attempts have also been made to explain why technological advancement may be biased—though it appears to be a much easier task to categorize the direction of technological bias than to explain it.

The purpose of this paper is to set forth a model which may be useful both in categorizing technological change and in testing for induced technological bias. The definitions of technological change and technological bias are set forth in the first section of the paper. Section II contains a brief review of alternative models of induced technological bias. In Section III, an empirical model is specified and the tests for induced and noninduced bias are set forth. Section IV contains empirical results based upon electric utility generation data and presents a brief discussion of the policy implications for that industry.

I. Definitions of Technological Change and Bias

Assume a production function

$$Q_{|\underline{Z},T}=f(\underline{X}) \tag{1}$$

where Q = output, $\underline{X}$ = a vector of inputs, $\underline{Z}$ = a vector of "state of nature" variables, and T = time. Technological change is defined as an inward movement in input space of the production-isoquant frontier. For a well-behaved production process, after technological advancement it will always be possible to produce the same level of output with at least one input set in which the quantity of each input is less than what would previously have been required. Or conversely, for a given set of inputs, technological advancement permits the production of a larger quantity of output than would have been otherwise attainable. Thus, on the production side, the rate of technological change (T_c) can be measured as

$$T_c=\left.\frac{\partial\, ln\, Q}{\partial T}\right|_{\underline{X},\underline{Z}} \tag{2}$$

presuming production takes place on the isoquant frontier.

Technological change can also be measured on the cost side of the production dual. For a firm at internal equilibrium, given factor-input prices and other state of nature constraints, technological change would permit the firm to produce the same level of output at a lower level of expenditure. Thus, on the cost side, the rate of technological advancement can be measured as

$$T_c=\left.\frac{\partial\, ln\, C}{\partial T}\right|_{Q,\underline{P},\underline{Z}} \tag{3}$$

where $\underline{P}$ is the set of factor input prices and $C = \sum_i P_i X_i$.[1] This is a measure of the rate of technological advancement which will be used in this paper.

Technological change may be biased both with regard to the factor inputs and with regard to the scale characteristics of the production process. With regard to factor-input biases, the definition of neutrality depends upon what is presumed to be

*Graduate School of Business, University of Wisconsin-Madison. This research was supported by the Graduate School of the University of Wisconsin-Madison and the National Science Foundation. I would like to thank Laurits Christensen, Thomas Cowing, and Frank Gollop for their comments and contributions to this analysis. I especially appreciate the valuable research assistance of William O'Neil, David Reifschneider, and Jeffrey Small.

[1]Since the measure in equation (1) is constrained intertemporally to the same input set and the measure in equation (2) is not, we would not necessarily expect the alternative technology advancement measures to be of the same magnitudes.

held constant. Hicks defined technological change as being neutral for a firm in "internal equilibrium" if the marginal rate of technological substitution between each pair of inputs is independent of technological change. Hicksian neutrality has been interpreted as implying that the shift of the production isoquant down the firm's expansion path will not alter factor proportions (X_i/X_j) or factor-cost shares $(P_iX_i/\Sigma_i P_iX_i)$.[2] In this paper, the standard of factor shares will be used to categorize technological change. Given the existence of technological advancement, the measure of input bias (I_b) is

$$(4) \qquad I_{bi} = \left. \frac{\partial S_i}{\partial T} \right|_{\underline{Q},\underline{P},\underline{Z}}$$

where S_i is the cost share of the ith input factor. A positive value of I_{bi} implies that technological change is relatively ith factor using, while $I_{bi}<0$ implies a relative ith factor-saving advancement and $I_{bi}=0$ implies neutrality.

Technological change may also be "biased" with respect to the return-to-scale characteristics of the production process. Such a bias would alter the range over which returns to scale of a given degree could be realized—and thus possibly alter the output level at which minimum average costs could be attained. Such changes would alter the degree of competition supportable in a given industry and thus have significant public policy ramifications.

With regard to cost functions, the scale measure is given by

$$(5) \qquad S_c|_{\underline{P},\underline{Z}} = \frac{\partial \ln C}{\partial \ln Q} = \frac{\partial C}{\partial Q}\frac{Q}{C}$$

[2]However, Charles Blackorby, C. A. K. Lovell, and Marie Thursby demonstrate a technological shift down the expansion path can simultaneously result in the maintenance of the marginal rate of technical substitution (*MRTS*) and nonvarying factor shares only if the production technology is homothetic. For a "neutral" *MRTS* shift along the expansion path of a firm operating with a nonhomothetic production process, factor shares would not remain constant.

where $S_c<1$ implies the existence of economies of scale, $S_c=1$ implies constant returns to scale, and $S_c>1$ implies diseconomies of scale. The measure of technological scale bias (TS_c) is given by

$$(6) \qquad TS_c = \left. \frac{\partial S_c}{\partial T} \right|_{\underline{Q},\underline{P},\underline{Z}}$$

Assuming the sign of TS_c is the same over the output range, $TS_c<0$ implies that minimum efficient firm size (*MES*) is increased, $TS_c=0$ implies no change in *MES* and $TS_c>0$ implies *MES* can be attained at a lower level of output.

II. Models of Induced Technological Bias

A. *Input Bias*

Several models of induced factor bias have been suggested in the literature. William Fellner suggests that producers, though atomistic with regard to input markets, perceive the macro effects of aggregated individual actions and thus have expectations as to shifts in relative prices over time. Expecting relative price shifts, and having available alternative innovation options, the producer follows a least-cost technological program to achieve minimum costs for a given level of output under future production conditions.

Syed Ahmad suggests the existence of an innovative possibility curve (*IPC*) which is the envelope of "...all alternative isoquants (representing a given output on various production functions) which the businessman expects to develop with the use of the available amount of innovating skill and time (assumed constant through the analysis)" (p. 348). The nature of the *IPC* frontier is a technological question. The selection of a production isoquant on the *IPC* frontier is the economic issue. As technology advances, the entire *IPC* in input space shifts inward toward the origin. As factor-input prices change over time, the firm selects a new production isoquant consistent with cost minimization. The end result is induced factor-input bias. Ahmad's analysis is based on actual factor-input price changes while

Fellner's is based on expected shifts in input prices.

Hans Binswanger (1974a,b) expands the analysis of induced factor bias through utilization of a research model which subsumes Ahmad's analysis as a special case. In Binswanger's model, which is based upon consideration of potential research payoffs, both factor biases and the rate of technological advancement depend upon the relative productivity and price of alternative lines of factor-saving research, the research budget constraint, and the present value of the stream of factor input costs. He states that a "...rise in the wage rate...tends to increase primarily the relatively labor-saving line of research. In general, this will lead to a more pronounced labor saving bias..." (1974b, p. 957). Also, since the initial factor input-output ratios affect the present value of factor costs, differences in the scale of operation could affect both the bias and rate of technological progress if the initial production process is nonhomothetic.

B. *Scale Biases*

Models relating to scale biases are not nearly as well developed as those relating to factor-input biases. Most of the literature relating scale and innovation has focused on the question of whether large scale firms innovate at a faster rate than small scale firms. The traditional Schumpeter-Galbraithian position is stated as follows: "Most of the cheap and simple innovations have, to put it bluntly, been made... Because development is costly, it follows that it can be carried on only by a firm that has the resources which are associated with considerable size" (John Kenneth Galbraith, pp. 86–87). Presuming a lag in the diffusion of new innovations, an apparent scale-biased shift of the industry's cost function could reflect the fact that large-scale firms operating with larger research budgets are able to reduce costs at a faster rate than small-scale firms.

While a larger scale of operation might be argued to be more promotive of technological advancement, certain technological advances might themselves be conducive to larger-scale operations—and hence be classified scale biasing in nature. Not all innovations, however, can be adopted by every firm in an industry. It is likely that large-scale firms might employ some production techniques not employed by, or available to, small-scale firms. If certain technological advances can only be utilized in larger-scale operations, then the rate at which costs decline over time for a given level of output would itself be a function of the scale of operation—or conversely, the rate at which the scale elasticities change over time may be a function of the scale of operations.

III. Empirical Model

The empirical model used in this paper to test for technological bias is a variant of the transcendental logarithmic cost function developed by Laurits Christensen, Lawrence Lau, and Dale Jorgenson (1971, 1973). Given the production function in equation (1), a dual cost function for a cost-minimizing firm would take the general form:

$$C=h(Q,\underline{P},\underline{Z},T) \tag{7}$$

where the variables are as defined above. Ignoring for the present the time variable, the traditional translog cost function specifies the *log* of costs as being equal to a second-order Taylor-series expansion in the logged arguments of the reduced-form cost function, or

$$\begin{aligned} ln\,C = {} & \alpha_o + \sum_i \alpha_i\, ln\,P_i \\ & + \left(\tfrac{1}{2}\right)\sum_i \sum_j \alpha_{ij}\, ln\,P_i\, ln\,P_j \\ & + \beta_1\, ln\,Q + \left(\tfrac{1}{2}\right)\beta_2 (ln\,Q)^2 + \sum_i \beta_{1i}\, ln\,Q\, ln\,P_i \\ & + \sum_k \gamma_k\, ln\,Z_k + \sum_i \sum_k \gamma_{ik}\, ln\,P_i\, ln\,Z_k \\ & + \sum_k \delta_k\, ln\,Z_k\, ln\,Q + \left(\tfrac{1}{2}\right)\sum_l \sum_k \delta_{lk}\, ln\,Z_l\, ln\,Z_k \end{aligned} \tag{8}$$

Symmetry is assumed (i.e., $\alpha_{ij}=\alpha_{ji}$ and $\gamma_{jk}=\gamma_{kj}$) and the following restrictions would reflect the condition of linear homogeneity in

prices:

$$(9)\qquad \sum_i \alpha_i = 1$$

$$\sum_i \alpha_{ij} = \sum_j \alpha_{ij} = \sum_i \sum_j \alpha_{ij} = 0$$

$$\sum_i \beta_{1i} = \sum_i \gamma_{ik} = 0$$

Using Shephard's Lemma, the ith input's cost share $(P_i X_i / C)$ is given by $(\partial \ln C / \partial \ln P_i)$. Also the scale elasticity is given by $(\partial \ln C / \partial \ln Q)$.

Viewing the time variable as a proxy for the level of technological development, we can incorporate time into the cost function to enable the measurement of the rate of technological change over a given time period. We incorporate time into the model by use of a truncated third-order Taylor-series expansion. A truncated third-order Taylor expansion is used instead of the normal translog second-order form for two reasons. First, one would expect that all coefficients of a normal translog function estimated on cross-sectional data would change as one moves from time period t to $t+j$. Under a normal translog format, when time is included as a "technology" variable, there is an implicit assumption that the non-time second-order coefficients do not change over time.[3] The truncated third-order model is more realistic in this regard, though at a cost of adding more coefficients to be estimated. Secondly (and perhaps more interestingly) the truncated third-order form allows us to specify tests for certain issues which would not be addressed under a normal translog formulation—specifically, we can test for price-induced technological input bias and we can test the Shumpeter-Galbraithian hypothesis that large firms have a more rapid rate of technological advancement than small firms. The intent in this regard is to show that the flexible functional form analysis can be "flexed" more than has traditionally been the case to enable an empirical examination of interesting economic issues.

[3]See, for example, Binswanger (1974a). His model is further constrained in that he assumes a homothetic production process.

The general model I propose to use is as follows:

$$(10)\qquad \ln C = H + \phi_1 T + \left(\tfrac{1}{2}\right)\phi_2 T^2$$

$$+ \sum_i \Psi_i T \ln P_i + \left(\tfrac{1}{2}\right)\sum_i \sum_j \Psi_{ij} T \ln P_i \ln P_j$$

$$+ \theta_1 T \ln Q + \left(\tfrac{1}{2}\right)\theta_2 T (\ln Q)^2$$

$$+ \sum_i \theta_{1i} T \ln Q \ln P_i + \sum_k \eta_k T \ln Z_k$$

$$+ \sum_i \sum_k \eta_{ik} T \ln P_i \ln Z_k + \sum_k \Omega_k T \ln Z_k \ln Q$$

$$+ \left(\tfrac{1}{2}\right)\sum_l \sum_k \Omega_{lk} T \ln Z_l \ln Z_k$$

where H represents the set of terms appearing on the right-hand side of equation (8). Equation (10) represents logged costs as a truncated third-order Taylor-series expansion in time and the logged input price, output, and state-of-nature variables. In order to conform with the assumption of linear homogeneity of prices, the restrictions set forth in (9) above and the following restrictions are required:

$$(9')\qquad \sum_i \Psi_i = \sum_i \Psi_{ij} = \sum_j \Psi_{ij} = \sum_i \sum_j \Psi_{ij} = 0$$

$$\sum_i \theta_{1i} = \sum_i \eta_{ik} = 0$$

The rate of technological change from equation (10) is given by

$$(11)\qquad \frac{\partial \ln C}{\partial T} = \phi_1 + \phi_2 T + \sum_i \Psi_i \ln P_i$$

$$+ \left(\tfrac{1}{2}\right)\sum_i \sum_j \Psi_{ij} \ln P_i \ln P_j$$

$$+ \theta_1 \ln Q + \left(\tfrac{1}{2}\right)\theta_2 (\ln Q)^2 + \sum_i \theta_{1i} \ln Q \ln P_i$$

$$+ \sum_k \eta_k \ln Z_k + \sum_i \sum_k \eta_{ik} \ln P_i \ln Z_k$$

$$+ \sum_k \Omega_k \ln Z_k \ln Q + \left(\tfrac{1}{2}\right)\sum_l \sum_k \Omega_{lk} \ln Z_l \ln Z_k$$

Factor-input-share bias, with regards to the ith input, is given by the expression

(12)
$$\frac{\partial S_i}{\partial T}=\Psi_i+\sum_j \Psi_{ij} \ln P_j+\theta_{1i} \ln Q+\sum_k \eta_{ik} \ln Z_k$$

Likewise, scale bias is given by the expression

(13)
$$\frac{\partial S_c}{\partial T}=\theta_1+\theta_2 \ln Q+\sum_i \theta_{1i} \ln P_i+\sum_k \Omega_k \ln Z_k$$

The extent to which factor-share bias is induced by factor-price shifts is given by

(14) $$\frac{\partial^2 S_i}{\partial T \partial P_j}=\Psi_{ij}$$

where we expect $\Psi_{ij}>0$ for $j \neq i$ and $\Psi_{ij}<0$ for $j=i$. In similar fashion, the extent to which scale bias is affected by output size is given by θ_2. Noting Binswanger's position with respect to the effects of initial factor input-output ratios on factor-share-biased technological change, the effect of scale on ith factor-share bias is given by θ_{1i}.

The translog cost equation and the associated derived factor-share equations are linear in their parameters and susceptible, upon making the necessary assumptions about the applicable stochastic error terms, to the application of classical least square regression techniques. Consequently, system estimation is not necessary to assure unbiasedness and consistency. However, as Jan Kmenta has noted, the existence of a high degree of multicollinearity among the exogenous variables "...is harmful in the sense that the estimates of the regression coefficients are highly imprecise. The implication arises because of the large variance of the least square estimators" (p. 389).[4]

As Christensen and William Greene indicate, "...the optimal procedure is to jointly estimate the cost function and the cost share equations as a multivariate regression system" (p. 662). Coefficient estimates can be derived by applying an adaption of Arnold Zellner's approach, developed for estimating seemingly unrelated equations. To apply Zellner's method to the set of translog cost and factor-share equations, however, one of the share equations must be eliminated to avoid singularity of the disturbance covariance matrix. The results of the estimating procedure, unfortunately, will not be invariant to the selection of the share equation designated for elimination. However, A. P. Barten has shown that the maximum likelihood estimator of a set of share equations with one equation deleted is invariant to the choice of the deleted equation. And, Kmenta and R. F. Gilbert have shown that an iteration of the Zellner procedure will converge to the maximum likelihood results.

The overall statistical significance of technological change may be measured by employing a likelihood ratio test. Defining L_R and L_U as the maximum likelihood values of the restricted and unrestricted model, respectively, note that the term $-2\ln(L_R/L_U)$ has a *chi*-square distribution with degrees of freedom equal to the number of independently imposed restrictions. The following restriction would be applicable for the likelihood ratio test, assuming equation (10) is the unrestricted model:

No Overall Technology Effect:

(15a) $$\phi=\Psi_i=\Psi_{ij}=\theta_1=\theta_2=\theta_{1i}=\eta_k=\eta_{ik}=\Omega_k=\Omega_{lk}=0 \quad (\text{for all } i,j,k,l)$$

No Factor-Input Bias:

(15b) $$\Psi_i=\Psi_{ij}=\theta_{1i}=\eta_{ik}=0 \quad (\text{for all } i,j,k)$$

No Scale Bias:

(15c) $$\theta_1=\theta_2=\theta_{1i}=\Omega_k=0 \quad (\text{for all } i,k)$$

No Price-Induced Factor-Share Bias:

(15d) $$\Psi_{ij}=0 \quad (\text{for all } i,j)$$

[4]The translog cost function has been criticized on the basis that an increase in the number of explanatory variables is likely to lead to a high degree of multicollinearity.

No Scale-Induced Factor-Share and Scale Bias:

$$\theta_{1i} = \theta_2 = 0 \qquad \text{(for all } i\text{)} \tag{15e}$$

IV. Technological Change in the Electric Utility Generation Industry

The technology model described above was estimated for the privately owned electric utility generating industry.[5] The model was estimated using firm level data for a pooled cross-section time-series sample consisting of eighty-one firms for the two years 1964 and 1972. Firms included in the sample generated at least 75 percent of their output from fossil fuel plants and operated with an average capacity utilization rate of at least 35 percent. Most of the data for the study is derived from the various issues of the Federal Power Commission's (*FPC*) *Statistics of Privately Owned Electric Utilities in the United States*, the *FPC*'s *Performance Profiles: 1963-70*, and J. Edward Smith's *The Measurement of Electric Utility Efficiency*. The "full technology" (equation (10)) cost function used for this analysis is[6]

$$\begin{aligned}
(16)\quad ln\,C &= \alpha_o + \sum_{i=1}^{3} \alpha_i \,ln\,P_i \\
&+ \left(\tfrac{1}{2}\right) \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_{ij}\,ln\,P_i\,ln\,P_j \\
&+ \beta_1\,ln\,Q + \left(\tfrac{1}{2}\right)\beta_2(ln\,Q)^2 \\
&+ \sum_{i=1}^{3} \beta_{1i}\,ln\,Q\,ln\,P_i \\
&+ \gamma_1\,ln\,CU + \left(\tfrac{1}{2}\right)\gamma_2(ln\,CU)^2 \\
&+ \sum_{i=1}^{3} \gamma_{1i}\,ln\,CU\,ln\,P_i + \delta\,ln\,CU\,ln\,Q \\
&+ \phi T + \sum_{i=1}^{3} \Psi_i T\,ln\,P_i \\
&+ \left(\tfrac{1}{2}\right) \sum_{i=1}^{3} \sum_{j=1}^{3} \Psi_{ij} T\,ln\,P_i\,ln\,P_j \\
&+ \theta_1 T\,ln\,Q + \left(\tfrac{1}{2}\right)\theta_2 T(ln\,Q)^2 \\
&+ \sum_{i=1}^{3} \theta_{1i} T\,ln\,Q\,ln\,P_i \\
&+ \eta_1 T\,ln\,CU + \left(\tfrac{1}{2}\right)\eta_2 T(ln\,CU)^2 \\
&+ \sum_{i=1}^{3} \eta_{1i} T\,ln\,CU\,ln\,P_i \\
&+ \Omega T\,ln\,CU\,ln\,Q
\end{aligned}$$

where the variables are defined as follows: C = costs, P_1 = fuel price, P_2 = labor wage, P_3 = capital price, Q = kilowatt hours generated, CU = average capacity utilization, and T = time. This equation will be referred to as Model 1. For estimation, the factor price, output, and capacity utilization variables have been normalized by the mean of their respective pooled-data vector.

Total costs are assumed to be equal to the summation of expenditures for fuel, labor, and imputed expenditures for capital services. Fuel expenditures and associated average fuel prices are obtained directly from the *FPC* and Smith data sources. Labor expenditures are calculated as the sum of salaries, wages, and employee pensions and benefits. Average yearly wage rates are estimated as the labor expenditures divided by the sum of the number of regular

[5]The technology model specified in Section III presumes cost minimization subject only to the production function constraint. The appropriateness of this assumption has been questioned by Harvey Averch and Leland Johnson (hereafter, A-J) and tentative support for the A-J hypothesis has been offered by Thomas Cowing, H. Craig Peterson, among others. William Boyes however offers contradictory evidence and Stephen Breyer and Paul MacAvoy (p. 108) indicate that the behavior of the electric utilities in the 1960's is contrary to the A-J hypothesis. Following Christensen and Greene the assumption of cost minimization will be maintained.

[6]The model is constructed for deriving parameter estimates from a two-year pooled cross-sectional data sample. Since time is to be treated as a binary variable, the term T^2 does not appear in equation (16). As a consequence, the model set forth in equation (16) does not permit the evaluation of changing rates of technological change over time for given levels of Q, $\underline{P}$, and Z.

full-time employees plus one-half of the number of part-time and temporary employees. Capital expenditures are calculated as the multiple of the deflated capital stock and the imputed capital service price. The deflated capital stock for an electric utility is computed as the sum of the deflated yearly net additions to plant service, correcting for mergers. The method used to compute a "price constant" capital stock for a utility is as follows:

$$(17) \qquad CS_i = CS_{i-1} + \frac{NI_i}{HW_i}$$

where

CS_i = adjusted capital stock for year i

NI = net investment for year i

HW_i = Handy-Whitman index for year i (adjusted to reflect a 1967 base of 100)[7]

$i = 1951\ldots,1972$

The base-period capital stock, CS_{1950}, is derived employing a triangularized weighted average of the Handy-Whitman index:

$$(17') \qquad CS_{1950} = \frac{K}{\sum_j^{20}\left[\left(j/\sum_j^{20}\right)HW_j\right]}$$

where j = the jth year from 1931 to 1950, and K = book value of plant in service in 1950. To take account of the acquisitions of plant and equipment through merger, all mergers among class A and B electric utilities are identified for the period 1950–72, and an adjusted capital stock measure is computed for each company which was ultimately acquired over the period of its corporate existence. The adjusted capital stock for any company, in year 1972 for example, is computed as the adjusted capital stock of that company net of acquisitions plus the adjusted capital stock of those class A and B companies which were acquired during the period 1950–72. Capital stock for generation plant is estimated as the ratio of the book values of generation plant and total plant in service times the deflated capital stock measure.

The imputed capital service price for a given utility is estimated for the ith year as

$$(18) \qquad P_{3i} = (OR_i + DE_i)HW_i$$

where OR = the utility's cost of capital, and DE represents the depreciation rate.[8]

Output is measured as kilowatt hours of net generation. Capacity utilization (output/[total kilowatt installed capacity x 8760 hours]) is included as a "state of nature" variable. Capacity utilization reflects both the load characteristics of the utility and the level of reserve margins.

Assuming symmetry and applying the linear homogeneity in price restrictions (see (9) and (9')) leaves thirty coefficients in equation (16) to be estimated. The full-cost function (Model 1) is estimated jointly with the fuel and labor cost share equations and the estimated coefficients are reported in Table 1.[9] Though the specified cost function (Model 1) is suitable for a nonhomothetic production process, we can test for homogeneity or homotheticity. If the production process is homogeneous (in both time periods) the following restrictions should be applied to the cost and associated factor-share equations:

$$(19) \qquad \beta_2 = \beta_{11} = \beta_{12} = \theta_2 = \theta_{11} = \theta_{12} = 0$$

Likewise, if the production process (in both time periods) is homothetic but not homogeneous, then the following restrictions would apply:

$$(19') \qquad \beta_{11} = \beta_{12} = \theta_{11} = \theta_{12} = 0$$

[7]The Handy-Whitman indices are a set of price indices for electric utility plants. The indices are reported for six geographic regions.

[8]The utility's cost of capital was estimated as a weighted average of its embedded cost of long-term debt, embedded cost of preferred stock, and the cost of common stock. The firm's capital structure was used to weight each capital component and a discounted cash flow model was used to estimate the financial cost of common stock capital.

[9]For all models, tests of monotonicity and concavity with respect to input prices passed. Symmetry was assumed.

TABLE 1

.Variable	Coefficient	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
One	α_0	18.145^{a}	18.167^{a}	18.158^{a}	18.068^{a}	18.092^{a}	18.137^{a}	18.145^{a}	18.145^{a}	18.154^{a}
$\ln P_1$	α_1	.381^{a}	.380^{a}	.380^{a}	.407^{a}	.408^{a}	.382^{a}	.380^{a}	.382^{a}	.381^{a}
$\ln P_2$	α_2	.090^{a}	.098^{a}	.098^{a}	.83^{a}	.083^{a}	.090^{a}	.092^{a}	.090^{a}	.093^{a}
$(\ln P_1)^2$	α_{11}	.152^{a}	.158^{a}	.151^{a}	.135^{a}	.136^{a}	.154^{a}	.142^{a}	.153^{a}	.137^{a}
$(\ln P_2)^2$	α_{22}	.069^{a}	.064^{a}	.062^{a}	.051^{a}	.056^{a}	.068^{a}	.049^{a}	.069^{a}	.046^{a}
$\ln P_1 \ln P_2$	α_{12}	−.025^{b}	−.030^{b}	−.027^{b}	−.012^{c}	−.015^{b}	−.025^{b}	−.013^{c}	−.025^{b}	−.010^{c}
$\ln Q$	β_1	1.021^{a}	.986^{a}	1.040^{a}	.932^{a}	.996^{a}	.996^{a}	1.021^{a}	1.012^{a}	1.005^{a}
$(\ln Q)^2$	β_2	.090^{a}		.088^{a}	.048^{b}	.089^{a}	.048^{b}	.092^{a}	.079^{a}	.078^{a}
$\ln Q \ln P_1$	β_{11}	.001			.014^{b}	.013^{b}	.003	.001	.003	.003
$\ln Q \ln P_2$	β_{12}	−.011			−.014^{a}	−.014^{a}	−.010^{a}	−.011^{a}	−.010^{a}	−.010^{a}
$\ln CU$	γ_1	−.763^{a}	−.740^{a}	−.796^{a}	−.747^{a}	−.700^{a}	−.833^{a}	−.760^{a}	−.752^{a}	−.837^{a}
$(\ln CU)^2$	γ_2	.459	.504	.480	−.166	.373	.414	.397	.454	−.035
$\ln CU \ln P_1$	γ_{11}	.156^{a}	.162^{a}	.158^{a}	.122^{a}	.126^{a}	.156^{a}	.152^{a}	.154^{a}	.130^{a}
$\ln CU \ln P_2$	γ_{12}	−.031^{c}	−.050^{a}	−.049^{c}	−.019^{c}	−.021^{c}	−.033^{c}	−.026^{c}	−.032^{c}	−.021^{c}
$\ln CU \ln Q$	δ	−.123	−.054	−.125^{a}	−.203^{a}	−.111	−.224^{a}	−.123	−.112	−.239^{a}
T	ϕ	−.116^{a}	−.120^{a}	−.128^{a}		−.030	−.102^{a}	−.115^{a}	−.119^{a}	−.149^{a}
$T \ln P_1$	Ψ_1	.046^{a}	.047^{a}	.046^{a}			.045^{a}	.047^{a}	.045^{a}	.045^{a}
$T \ln P_2$	Ψ_2	−.014^{b}	−.023^{a}	−.023^{a}			−.015^{a}	−.015^{a}	−.014^{a}	−.016^{a}
$T(\ln P_1)^2$	Ψ_{11}	−.013	−.014	−.012			−.018		−.014	
$T(\ln P_2)^2$	Ψ_{22}	−.034^{c}	−.033^{c}	−.030^{c}			−.035^{c}		−.033^{c}	
$T \ln P_1 \ln P_2$	Ψ_{12}	.017	.017	−016			.020^{c}		.018	
$T \ln Q$	θ_1	−.082^{b}	−.024	−.077^{b}		−.076^{b}		−.083^{b}	−.072^{a}	−.068^{b}
$T(\ln Q)^2$	θ_2	−.023		−.024		−.015		−.025		
$T \ln Q \ln P_1$	θ_{11}	.004						.004		
$T \ln Q \ln P_2$	θ_{12}	.002						.001		
$T \ln CU$	η_1	.011	−.009	.046		−.079	.074	.010	−.0002	−.068
$T(\ln CU)^2$	η_2	−2.242^{b}	−2.072^{b}	2.246^{b}		−2.22^{b}	−2.166^{b}	−2.238^{b}	−2.269^{b}	
$T \ln CU \ln P_1$	η_{11}	.040	−.045	−.042			−.043	−.036	−.037	
$T \ln CU \ln P_2$	η_{12}	.012	.030	.020				.009	.142	
$T \ln CU \ln Q$	Ω	−.275^{b}	−.292^{b}	−.276^{c}		−.312^{b}	.016	−.289^{b}	−.296^{b}	
Logged Likelihood Value		750.722	736.799	742.597	725.984	736.382	744.389	749.175	750.423	742.665

$^{a}|t| > 2.576$
$^{b}|t| > 1.96$
$^{c}|t| > 1.28$

Cost and factor-share functions incorporating the respective restrictions of (19) and (19′) are estimated in a similar fashion as Model 1. The two restricted cost functions are labeled Model 2 (homogeneity) and Model 3 (homotheticity) and the estimated coefficients for the two models are reported in Table 1. Casting Model 1 as the unrestrained model and Models 2 and 3 as the restrained models, we can test the joint hypotheses set forth in (19) and (19′). As indicated in Table 2, these hypotheses are not sustainable. Thus, we conclude that it is appropriate to construct the cost function in such a manner as to be consistent with an underlying nonhomothetic production process.

Restricted versions of Model 1 are estimated to test the various joint hypotheses with respect to technological impact, factor-input bias, scale bias, factor-price-induced effects, and scale-induced effects (see equations (15)). These results are also set forth in Table 1. Model 1 is the full-technology model; Model 4 assumes no technological advancement effect; Model 5 assumes a technological effect but no factor-share bias; Model 6 assumes a technological effect but no scale bias; Model 7 assumes a technological effect but no price-induced factor-share biasing effect; and Model 8 assumes a technological effect but no scale-induced factor-share or scale elasticity bias. Taking Model 1 as the unrestricted model

and Models 4–8 as the restricted models, each restricted model is tested using the likelihood ratio test. The results of those tests are set forth in Table 2. As indicated in Table 2, we can reject the hypotheses of no technological effect, no factor-cost-share biasing effect, and no scale elasticity biasing effect. However, we cannot reject the hypothesis of no price- or scale-induced biases. As noted in Model 8, however, the signs of the price-induced coefficients are all in the right direction. One additional model was tested. Model 9 assumed that the cost function could be represented as a second-order Taylor-series expansion in time and the logged variables P_1, P_2, P_3, Q, and CU—as opposed to Model 1 which is a truncated third-order expansion. As indicated in Table 2, we cannot accept the null hypothesis that Model 9 provides as good explanatory powers as Model 1.

Given the estimated technology model (Model 1), we can evaluate the extent of technological advancement and any associated factor or scale bias.[10] For a firm operating both in 1964 and 1972 at the expansion point of our data sample (i.e., operating with an output, capacity utilization level, and factor prices equal to the mean of the sample's respective pooled-data vectors), the estimated effect of technological advancement over the period would have been to reduce the cost of production by 11.6 percent. The estimated factor-biasing effect of the technological change, for the firm operating at the expansion point, would have been to increase the fuel cost share from 38.1 to 42.7 percent, reduce the labor cost share from 9.0 to 7.6 percent, and reduce the capital cost share from 52.9 to 49.7 percent. At the expansion point, estimated scale economies in 1964 would have been 1.021 (with a standard error of .03). In 1972, however, if the firm continued to operate at the expansion point, the estimated scale-biasing effect of technological change would have been to reduce the scale elasticity estimate to .994 (with a standard error of .05).

TABLE 2

Model	$-2ln(L_R/L_U)$	Restrictions	Significance
2	27.846	6	.005
3	16.250	4	.005
4	49.476	15	.005
5	28.680	9	.005
6	12.666	5	.05
7	3.094	3	NS
8	.598	3	NS
9	16.114	10	.1

Notes: Model 1 is the unrestricted model (L_U). Models 2–9 are the restricted models (L_R). NS=No significant difference w/r to Model 1.

An alternative way to evaluate the estimated nature of technological change is to consider the observational values for $[\partial ln C/\partial T]$ in equation (11); $[\partial S_i/\partial T]$ in equation (12); and $[\partial S_c/\partial T]$ in equation (13) over the range of the data in our sample. With respect to our pooled 1964 and 1972 data, the average estimated rate of technological advancement is 13.8 percent for the eight-year span with a standard error of dispersement of 8.9.[11] With respect to the pooled data, $\partial ln C/\partial T$ is negative for seventy-nine of eighty-one 1964 observations and eighty of the eighty-one 1972 observations. With respect to factor-share biasing, for all 1964 and 1972 observations $\partial S_i/\partial T$ is positive for the fuel share and negative for the capital share. For the labor share $\partial S_i/\partial T$ is negative for all but one 1972 observation. The respective averages of $\partial S_i/\partial T$ when evaluated at each of the 1964

[10]For Model 1 the estimated price and scale elasticities for the 1964 and 1972 observational values are all positive. For $\partial ln C/\partial ln P_i$ the means and standard error of dispersement are .396 (.065) for fuel, .090 (.034) for labor, and .514 (.092) for capital. The average observational value for $\partial ln C/\partial ln Q$ is .95 with a standard error of dispersement of .077.

[11]The standard error of dispersement is based on the range of the estimated observational values of $\partial ln C/\partial T$, etc., and is not to be considered as the standard error of the estimate. The standard error of $\partial ln C/\partial T$, etc. at any observation point can be calculated with information on the variance-covariance matrix and by noting that if the error terms of the cost and factor-share equations have a joint normal distribution, $\partial ln C/\partial T$, etc. are linear combinations of normally distributed variables.

and 1972 observations (and the associated standard errors of dispersement) are .046 (.008) for fuel, −.036 (.011) for capital, and −.015 (.008) for labor. With regard to the estimated effect of technological advancement on scale elasticities, $\partial S_c/\partial T$ evaluated over the data is negative for sixty-nine of the eighty-one 1964 observations and eighty of the eighty-one 1972 observations. The average of the $\partial S_c/\partial T$ observational values, with regard to the data, is −.070 with a standard error of dispersement of .054.

Several policy conclusions might be drawn from the technology model with respect to the electric utility generating industry. Many writers have suggested that the inherent scale economies in the electric utility generating industry could best be realized through increased system coordination accomplished by merger or other means (note William Hughes; Breyer and MacAvoy). Donald C. Cook, the former chairman of American Electric Power Company (the largest privately owned electric utility in the United States), has suggested that all electric utilities should be merged into six to twelve regional megasystems. Christensen and Greene's analysis of scale economies, however, does not support the massive proposals. The Christensen and Greene results, while evidencing the potential desirability of merger among very small electric utilities, indicate that minimum efficient firm size would be reached at an output level substantially below that which would occur if the Cook merger proposals were followed. However, it might be argued that merger among large firms would still be desirable if it could be demonstrated that larger firms innovate at a significantly higher rate than smaller firms. The results of this paper do not substantiate the Schumpeter-Galbraithian hypothesis that large firms innovate at a more rapid pace than small firms. If the Schumpeter-Galbraithian hypothesis held, we would expect $\partial^2 ln C/\partial T \partial ln Q$ to be a decreasing function of Q, or $ln\ Q$. Though the value of $\partial(\partial Sc/\partial T)/\partial\ ln\ Q = \theta_2 = -.023$, the coefficient, θ_2, can not be shown to be significantly different from zero. We should also note that our estimated cost model can be shown to support Christensen and Greene's earlier findings with respect to scale economies.

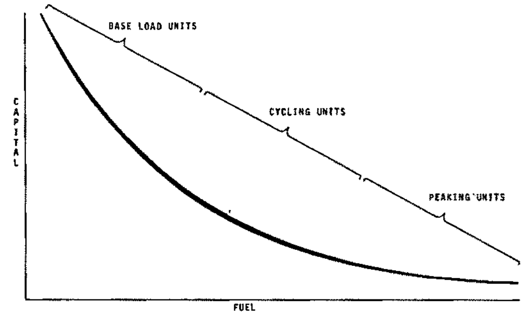


FIGURE 1

Another policy conclusion which might be drawn from this analysis pertains to the advantages of "load flattening" which may be the result of "time of day" pricing or other peak load pricing variants. Systems which have flatter loads (i.e., a more constant usage pattern) tend to operate with a higher proportion of base load plants to cycling to peaker units. Base load units have a higher capital fuel ratio compared to the cycling and peaker units. Assuming a constant level of labor input, the isoquant curve of generating units can be described as in Figure 1.

The nature of innovation described by Model 1 indicates that the technological advancement has been relatively capital saving and fuel using. At constant factor-input prices, this would imply that a cost-minimizing firm producing a given level of output would be induced to substitute capital for fuel in the production process. Since the plant mix of a utility is in part determined by the shape of the firms's load curve, the substitution of more capital intensive (for example, base load) units for fuel intensive (for example, peaking) units is facilitated if the firm's load curve can be flattened. While peak load pricing was most likely justified in 1964, the nature of the subsequent technological advancements have increased the advantages which would stem from this type of rate structure reform in the electric utility industry.

V. Conclusions

In this paper, a model based on the translog cost function and designed to test for technological advancement, factor-input and scale biases arising from technological advancement, and induced technological bias was set forth. The model was estimated with data from the electrical utility generation industry. While the empirical results for that industry demonstrated the existence of technological advancement and technological bias, both with regard to factor inputs and scale economies, the results failed to demonstrate the existence of induced technological bias.

One further observation should be made. Previous studies of technological advancement in the electric utility industry have shown yearly total factor productivity advancements of 2.5 to 5 percent (see Kendrick; Yoram Barzel). These studies generally covered an earlier time span than considered in this study. At the expansion point, the results of this study indicate an average annual rate of technological advancement of approximately 1.5 percent while the average "observed" rate is 1.7 percent. Two conclusions could be drawn: either the rate of technological advancement has declined substantially over time, or the total factor productivity measures developed in the previous studies overstate technological advancement due to a failure to adjust for the effects of economies of scale and capacity utilization or due to errors in measurement.

REFERENCES

S. Ahmad, "On the Theory of Induced Invention," *Econ. J.*, June 1966, *76*, 344–57.

H. Averch and L. L. Johnson, "Behavior of the Firm under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

A. P. Barten, "Maximum Likelihood Estimation of A Complete System of Demand Equations," *Euro. Econ. Rev.*, Fall 1969, *1*, 7–73.

Y. Barzel, "Productivity in the Electric Power Industry," *Rev. Econ. Statist.*, Nov. 1963, *45*, 395–408.

H. P. Binswanger, (1974a) "The Measurement of Technical Change Biases with Many Factors of Production," *Amer. Econ. Rev.*, Dec. 1974, *64*, 964–76.

______, (1974b) "A Microeconomic Approach to Induced Innovation," *Econ. J.*, Dec. 1974, *64*, 940–58.

C. Blackorby, C. A. K. Lovell, and M. C. Thursby, "Extended Hicks Neutral Technical Change," *Econ. J.*, Dec. 1976, *66*, 845–52.

W. G. Boyes, "An Empirical Examination of the Averch-Johnson Effect," *Econ. Inquiry*, Mar. 1976, *14*, 25–35.

Stephen G. Breyer and Paul W. MacAvoy, *Energy Regulation by the Federal Power Commission*, Washington 1974.

L. W. Christensen and W. H. Greene, "Economies of Scale in U.S. Electric Power Generation," *J. Polit. Econ.*, Aug. 1976, *84*, 665–76.

______, **D. W. Jorgenson, and L. J. Lau,** "Conjugate Duality and the Transcendental Logarithmic Production Function," *Econometrica*, July 1971, *39*, 255–56.

______, ______, **and** ______, "Transcendental Logarithmic Production Frontiers," *Rev. Econ. Statist.*, Feb 1973, *55*, 28–45.

T. G. Cowing, "The Effectiveness of Rate of Return Regulation: An Empirical Test Using Profit Functions," in Melvin Fuss and Daniel McFadden, eds., *Production Economics: A Dual Approach to Theory and Application*, Amsterdam 1977.

W. Fellner, "Two Propositions in the Theory of Induced Innovations," *Econ. J.*, June 1961, *71*, 305–08.

John Kenneth Galbraith, *American Capitalism*, Boston 1956.

J. R. Hicks, *The Theory of Wages*, New York 1932.

W. R. Hughes, "Scale Frontiers in Electric Power," in William M. Capron, ed., *Technological Change in Regulated Industries*, Washington 1971.

John W. Kendrick, *Productivity Trends in the United States*, Princeton 1961.

Jan Kmenta, *Elements of Econometrics*, New York 1971.

______ **and R. F. Gilbert,** "Small Sample Properties of Alternative Estimators of

Seemingly Unrelated Regression," *J. Amer. Statist. Assn.*, Dec. 1968, *58*, 1180–200.

H. C. Peterson, "An Empirical Test of Regulatory Effects," *Bell J. Econ.*, Spring 1975, *6*, 111–26.

Ronald W. Shephard, *Cost and Production Functions*, Princeton 1953.

J. Edward Smith, *The Measurement of Electric Utility Efficiency*, Washington 1976.

A. Zellner, "AnEfficient Methof For Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Assn.*, June 1962, *57*, 585–612.

Federal Power Commission, (*FPC*) *Statistics of Privately Owned Electric Utilities in the United States*, Washington, various years.

______, *Performance Profiles: 1963–1970*, Washington 1973.

***Handy-Whitman Indexes of Utility Construction Costs*,** Whitman, Requardt, and Associates, 1977.

Queuing Analysis and Value of Service Pricing in the Trucking Industry: Comment

By KENNETH D. BOYER*

In a recent issue of this *Review*, A. S. De Vany and T. R. Saving (hereafter, D-S) presented a new perspective on pricing in the trucking industry. The authors' principal claim is that with their approach to service "quality," value-of-service pricing could be expected in a perfectly competitive trucking industry. They contrasted their results with work presented earlier by Josephine Olson. She used measurements of value-of-service elements in a motor carrier class rate structure to infer that the industry engaged in monopoly pricing. The present paper argues that an analysis of queuing aspects of quality is inappropriate for the trucking industry.

There are two commonly used meanings for value-of-service pricing. It can refer either to the general practice of price discrimination by freight carriers (see George Wilson), or to the specific discriminatory practice of varying charges for the same service according to the dollar value of the commodity carried (see John Meyer et al.). (These two definitions are neither identical or necessarily closely related if there are many modes and multiple markets.) Since value-of-service pricing has been seen as discriminatory, it has been associated with market power in the transport industries. Section I of this paper points out that the value-of-service pricing described by D-S is not discriminatory; thus the perception of discrimination requiring market power is not contradicted.

Succeeding sections argue that there is a logical inconsistency in D-S's paper when they argue that their results are insensitive to the assumptions on risk pooling; that competitive firms who carry more than one commodity or who operate on more than one haul would not engage in value-of-service pricing. It is further argued that in fact we do not observe the type of behavior predicted by a queuing analysis of motor carrier service quality. The final section shows that, even if their analysis were flawless, the price variations predicted by D-S's model are several orders of magnitude smaller than those actually observed.

*Associate professor of economics, Michigan State University. This research was funded by a grant from the ATA Foundation.

I. Product Quality and Value-of-Service Pricing

William Baumol and David Bradford claimed that value-of-service pricing could be properly seen as a second best position if natural monopolies were subject to a break-even constraint. De Vany and Saving, however, argue that value-of-service pricing might represent a first best social optimum.

To reach this conclusion, the authors put forth a model of the trucking industry with queuing. The industry is composed of a series of firms which carry truckloads of a single commodity between two end points. With stochastic demand for service, a limited number of vehicles in each market leads to the random growth of queues. Demand for a motor carrier's service then depends on both the price charged and the length of the wait expected by a shipper. Prices are invariant to queue lengths, so that the distribution of customers among carriers is achieved not through price rationing, but through balking at excessive waiting time. Capacity also is fixed with respect to queue lengths. In particular, a carrier cannot hire vehicles quickly enough to prevent the random development of queues whose length leads shippers to balk and turn to other carriers.

With uncontrollable queues of random length below a balking maximum, the

steady state for an industry is characterized by a positive expected waiting time for each shipper. With costly inventory, a shipper must expect to pay a "full price" (shipping charges plus inventory holding costs) for freight services. Industry capacity adjusts to provide an expected queue length which minimizes social costs. If the commodity shipped has a high value, the socially optimal wait time will be short; low-valued commodities, however, have low carrying costs and hence shippers will tolerate the longer queues associated with less "excess capacity" in the industry.[1] Since entry drives expected economic profits to zero, less excess capacity should lead to lower charges. Since it is socially optimal to provide shippers of valuable merchandise with a higher quality service, the shippers of high-value goods will pay for the higher quality service in higher rates. Thus D-S claim that imposing higher rates on higher-value goods is both socially optimal and consistent with perfect competition in the trucking industry.

It should be noticed, however, that the authors have not described the discriminatory value-of-service pricing criticized as yielding an excess burden. De Vany and Saving present a model in which a higher-cost service is provided at a higher price to shippers of high-value merchandise. Thus what appears to be price discrimination actually is not. A careful reading makes it clear that the queuing analysis does not claim that discrimination is socially optimal. When costs are properly calculated, the competitive firms do price at marginal cost.

Can we then use the queuing approach to product quality to demonstrate that value-of-service pricing as currently practiced is nondiscriminatory and a first best social optimum? The following three sections of this paper argue that we cannot.

[1]Since a fixed number of truckloads are shipped per period, albeit at a random rate, D-S are able to define excess capacity as using more trucks than would be physically required to haul the fixed number of tons at a constant rate. This is a different definition from the orthodox idea of producing in the downward-sloping portion of the *LRATC* curve for a constant quality product.

II. Queuing Requires Single Market Firms

In D-S's analysis, all trucking firms operate in single markets (or exactly two markets, counting the backhaul as separate). Markets are defined as the transport of a single commodity between two end points. They note that if firms operate in several markets, there will be benefits to risk pooling, but dismiss risk pooling in a footnote as not of interest in their paper. In fact, assuming away risk pooling is the key to the outcome of their analysis.

De Vany and Saving present their fundamental theorem in equations (39) and (40):

$$p_A^i = C_{\alpha i} + \alpha^i \eta_A \frac{\partial \overline{W}_A}{\partial \alpha^i} - \beta^i \eta_B \frac{\partial \overline{W}_B}{\partial \alpha^i} \tag{39}$$

$$p_B^i = C_{\beta i} + \beta^i \eta^B \frac{\partial \overline{W}_B}{\partial \beta^i} \tag{40}$$

Equation (40) states that a competitive firm in equilibrium will charge a price for the backhaul equal to the movement costs plus a congestion toll. This toll is equal to the carrying cost per truckload of inventory η^B, multiplied by the number of shipments on the backhaul β^i, multiplied by the increase in the average elapsed time between pickup and delivery for a one-truckload increase in shipments on the backhaul $\partial \overline{W}_B / \partial \beta_i$. Equation (39), for the fronthaul, includes congestion tolls both for the front and backhauls.

It is the presence of η^A and η^B in equations (39) and (40) on which D-S base their arguments concerning value-of-service pricing. Their competitive rates depend on the carrying cost of inventory, which in turn is presumed related to the value of merchandise carried. Higher inventory carrying costs induce truckers to have more capacity devoted to hauling this good—capacity which stands idle more of the time than that devoted to low-value goods. The congestion tolls in (39) and (40) are charges to offset expected idle capacity of hauling this commodity. The extra terms in (39) and (40) are there only because capacity is assumed to

be specific to the haulage of single commodities.

Equations (39) and (40) can be clearly interpreted only if capacity is dedicated to hauling a single commodity between two end points. But by engaging in such market dedication of capacity, a firm forgoes the benefits of risk pooling. If firms can shift equipment from hauls and commodities where demand is momentarily low to markets where demand is unusually heavy, and if firms can meet peak periods by short-term equipment leasing, there is no reason for queues to develop and no reason to lose shipments due to customer balking.

It is only because they have assumed away equipment leasing and risk pooling that D-S obtain their striking results. This can be seen by examining their equations (39) and (40). Derivatives of $\overline{W}$, the average elapsed time between order and completion of service, enter both equations.

By their footnote 18,

$$\overline{W}=\Pi/(N\mu-\alpha) \tag{1}$$

where α = realized shipments per unit time, $N\mu$ = capacity, and Π is the probability that all capacity is busy. Following D-S, Π is treated as a constant.

$$\frac{\partial \overline{W}}{\partial \alpha}=\frac{-\Pi}{(N\mu-\alpha)^2}\left(-1+\frac{\partial(N\mu)}{\partial \alpha}\right) \tag{2}$$

If capacity can be assigned or leased as quickly as orders arrive, then $\partial(N\mu)/\partial\alpha=1$ and the right-hand side of equation (2) vanishes. So do the congestion tolls in (30) and (40). In fact, equations (39) and (40) can not be written since they are derived using the assumption that $\partial\overline{W}/\partial\alpha\neq 0$. If the derivative is zero, then there is no distinction between being a "transport-price taker" and a "full-price taker"; there is no queuing, no balking and thus there is a constant quality of service for all commodities.

III. The D-S Model does not Accord with Observation

DeVany and Saving's model assumes that equipment is dedicated to single commodities. It predicts that there will be a lower utilization rate for equipment devoted to carrying high-value commodities; these differential use rates will be reflected in higher charges and lower average service times for high-value goods.

This section argues that: 1) it is the rare commodity which has capacity devoted solely to handling it, and that in fact capacity is freely interchangeable among markets; 2) since capacity is not devoted to single commodities, the prediction of differential utilization rates can have no empirical content; 3) the corollary to differential utilization rates—faster average service times for high-value goods—is empirically contradicted. Thus the final assertion, that value-of-service pricing is a payment for different service qualities, must also be rejected.

A. *Trucking Capacity is not Commodity Specific*

Some groups of commodities do need special equipment. Notable are heavy machinery, refrigerated products, liquid petroleum products, and motor vehicles. These groups combined account for approximately 17 percent of revenues of the for-hire trucking industry. With the exception of household goods movers (4 percent of revenues), almost all other commodities are shipped in standard highway trailers.[2] These trailers are used to handle virtually any nonbulk commodity.

The Interstate Commerce Commission reports that on typical interstate highway segments which it checked during a one-year period, 50 percent of the trucks were standard vans, 16 percent were refrigerated vans, 18 percent were flats or lowboy vans, 8 percent were tank trucks, 3 percent were bulk carriers, and 4 percent were other types of trucks.[3] Of these, only tank trucks are likely to be dedicated to a single commodity. Even refrigerated vans can carry commodities with different inventory costs.

[2]Numbers are from American Trucking Associations (1977, p. 25).

[3]Numbers are from Interstate Commerce Commission, p. 12.

The largest number of shipments handled by the regulated motor carrier industry are *LTL* (Less Than Truckload).[4] As D-S note, *LTL* shipments are usually consolidated at terminals before the line-haul segment. What they fail to point out is that different goods are typically consolidated in the same truck. It is impossible to assign responsibility for idle *LTL* equipment to any commodity.

Additional evidence is provided by a survey of truckers I performed in July 1978. (See the Appendix for further details.) To determine if high- and low-value goods were shipped with the same equipment, the carriers were asked: "Would you use the same truck to ship either tires or newsprint?" One hundred percent of respondents answered affirmatively.

B. *The Question of Differential Utilization Rates Cannot be Answered*

Since capacity in the trucking industry is, with rare exception, freely interchangable among different commodities, the question of how use rates vary with commodity is meaningless. No data exist on utilization rates of capacity devoted to tire haulage as opposed to that used for newsprint service since such capacity does not exist.

C. *Service Quality is Invariant to Commodity*

There remains the logical possibility that, for reasons other than differential use rates, high-value goods on average receive prompter service than low-value goods. One reason might be that carriers regularly give priority to high-value shipments. There is, however, no evidence that carriers follow such a practice.

There are no published data on the average length of queues in trucking or on the frequency of balking. One suspects that the reason relates to the nonoccurence of such phenomena. There are similarly no published data on the time between order of transport services and receipt by consignee of different commodities. Since there are surveys of truck transit time,[5] but commodity designation is not included as an influencing factor, one is led to believe that order cycles are of the same length regardless of the good shipped.

In the literature of motor carrier management, reference to methods of queue management to achieve an optimum balking level cannot be found. What is present are discussions of optimum equipment assignment and dispatching (for example, see Charles Taff). One finds no discussion of how to place shipments in the dispatch queue according to commodity value. One can find a discussion on leasing owner-operators to take care of periods of peak demand (see Daryl Wycoff and David Maister).

In a recent survey, traffic managers were asked to list criteria used to select motor carriers; only 5.9 percent listed availability of standard equipment (see Kenneth Flood, p. 19). In a 1976 discussion in the trade press, among the reasons cited for switching to private carriage, the avoidance of equipment shortages does not appear.[6] In discussions of the factors determining the speed of truck delivery of a shipment, the commodity type is notably missing.[7]

It is somewhat disturbing that no reference can be found which proves the nonexistence of queuing and differential service speeds; still it is perhaps not surprising that along with the absence of references in the trade literature to the existence of queuing, balking, and differential service times, I similarly can find no statements that such phenomena do not exist. Should we expect, however, the trade literature to discuss the nonexistence of nonexistent phenomena? On the other hand, it would be astonishing indeed to have the phenomena of queuing and differential service speed ignored in the literature on the motor carrier industry if

[4]This can be confirmed by inspection of any of the yearly continuing traffic studies produced by Motor Carrier Rate Bureaus.

[5]See, for example, Keene Paterson and Herschel Cutler.

[6]See *Handling and Shipping*.

[7]The discussion of factors affecting speed of delivery includes the number of items shipped, the efficiency of loading docks, and traffic congestion, but not the commodity carried. See Donald Bowersox et al.

they played the central role that D-S claim for them.

D. *A Survey of Motor Carriers*

To overcome the lack of published data, I performed my own survey of motor carriers to determine whether the predictions of D-S's model are empirically validated. The survey was administered in July 1978 to all trucking firms having terminals in Lansing, Michigan. Of a total of fourteen firms, twelve responses were received.

All firms responding claimed that the average length of time from the moment shippers tell the carrier they have loads to when it is picked up was less than one day; the times reported varied from two hours to overnight, and all insisted that there was little variation in these figures.

In response to the question, "Does the average time between placing and pickup of a shipment depend on the commodity carried?," all twelve firms responded negatively. When asked, "How frequently do shippers turn to another carrier because you are short of equipment?," the responses were: once per year (2); rarely (6); infrequently (1); not often (1); occasionally (1); and very seldom (1).

The final question was: "If you have an equipment shortage, how do you determine which goods to ship first?" Eleven responses were on the general theme of first come, first served, or first protect regular shippers or shipments. One firm claimed never to have had the problem.

Surveys are not an ideal device for measurement and the survey just described is admittedly small. Nevertheless, to the extent that trucking operations in Lansing, Michigan are typical of those in the rest of the country, and to the extent that self-conscious answers can be believed, one must conclude that queuing does not exist in the trucking industry and that different commodities do receive the same speed of service.

IV. A Queuing Model Would Explain Only a Small Part of Value-of-Service Pricing

The basic assumption on which D-S's paper is based—no risk pooling—is invalid. Neither is there evidence to support the predictions of the model—queuing, balking, and variations in capacity utilization rates and average service times according to commodity value.

However, even if the model were correct in claiming that prices are used to reward commodity-specific queue-shortening capacity additions, queue rationing cannot explain observed pricing variations.

Table 1 displays the congestion toll markups implied by D-S's equation (40) under these assumptions: inventory carrying costs are 25 percent per year; an average of 100 tons per day are shipped 800 miles; truckloads are 24 tons. Numbers in Table 1

TABLE 1—VALUES OF CONGESTION TOLL

$$\beta^i \eta_B \frac{\partial \overline{W}_B}{\partial \beta^i}$$

Value of Commodity	$\frac{\partial \overline{W}_B}{\partial \beta^i}$ = 1 hr./truckload	$\frac{\partial \overline{W}_B}{\partial \beta^i}$ = 5 hrs./truckload	$\frac{\partial \overline{W}_B}{\partial \beta^i}$ = 24 hrs./truckload
\$200/ton	.00296 (.0318)	.00742 (.0897)	.0712 (.765)
\$500/ton	.00742 (.0897)	.0371 (.399)	.178 (1.91)
\$1000/ton	.01484 (.159)	.0742 (.897)	.356 (3.82)
\$2000/ton	.0296 (.318)	.148 (1.59)	.712 (7.65)
\$5000/ton	.0742 (.897)	.371 (3.99)	1.78 (19.1)

Notes: Congestion tolls are in cents per ton mile. Parenthetical figures are percentages of the average truck charge of 9.6 cents per ton mile.

are quoted in cents per ton mile and in percentages of the average *U.S.* motor carrier charge of 9.6 cents per ton mile.[8]

Price variations which have traditionally been attributed to value-of-service pricing are measured in hundreds of percents. For example, class ratings on different commodities vary by more than 400 percent. By contrast, the markups in Table 1 are in most cases less than 1 percent of the average truck charge. To explain a markup of 19 percent of the average charge using D-S's model, one would have to assume (a) very high value for the goods, and (b) that the offering of an additional truckload would increase average expected delivery time of all truckloads by 24 hours. In order to explain actual price variations, the derivative $\partial \overline{W}/\partial \alpha$ would have to be measured in weeks per truckload. This is totally implausible.

Thus, even were D-S's model correct, it could not explain the price variation that has traditionally been attributed to value-of-service pricing.

Appendix

In July 1978, the following questions were sent to all trucking firms listed as having terminals in Lansing, Michigan. Responses were received from the following twelve firms: Associated Truck Lines, Blue Arrow-Douglas, Branch Motor Express, Central Transport, Ogden and Moffett, McLean Trucking, Renner's Express, Roadway Express, Tucker Freight Lines, United Trucking, U.S. Truck Co., and Yellow Freight System. Copies of the complete responses are available from me on request.

1. How long, on average, is the time between when a shipper tells you he has an order and when you pick up the shipment? *Responses*: Two-three hours (2); afternoon if ordered in the morning, morning if ordered in the afternoon (3); same day (7).

2. Does this time vary a lot or is it a pretty reliable figure? *Responses*: Reliable (12).

3. Does the average time between placing and pickup of a shipment depend on the commodity carried? That is, on average, would you pick up a shipment of newsprint as quickly as an order of tires, or would you give one of them priority? *Responses*: Same service to both (12).

4. Would you use the same truck to ship either tires or newsprint? *Responses*: Same truck for both (12).

5. How frequently do shippers turn to another carrier because you are short of equipment? *Responses*: Once per year (2); not often, rarely, infrequently, occasionally or very seldom (10).

6. If you do have an equipment shortage, how do you determine which goods to ship first? *Responses*: First come, first served, protect regular customers first, or protect regular shipments first (11); never have the problem (1).

[8] This number is from American Trucking Associations (1976, p. 5).

References

W. F. Baumol and D. F. Bradford, "Optimal Departures From Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *60*, 265–83.

Donald J. Bowersox, Edward W. Smykay, and Bernard J. LaLonde, *Physical Distribution Management*, London 1969.

A. S. De Vany and T. R. Saving, "Product Quality, Uncertainty and Regulation: The Trucking Industry," *Amer. Econ. Rev.*, Sept. 1977, *67*, 583–94.

Kenneth Flood, *Traffic Management*, Dubuque 1975.

John R. Meyer et al., *The Economics of Competition in the Transportation Industries*, Cambridge, Mass. 1960.

J. E. Olson, "Price Discrimination by Regulated Motor Carriers," *Amer. Econ. Rev.*, June 1972, *62*, 395–402.

K. Paterson and H. Cutler, "Transportation and MILSTRIP," in Amer. Transportation Res. Forum, *1961 Papers of the 2d Annual Meeting*, New York 1961, pp. 1X-1–1X-17.

Charles Taff, *Commercial Motor Transportation*, Ithaca 1975.

George W. Wilson, *Essays on Some Unsettled Questions in the Economics of Transportation*, Bloomington 1962.

Daryl Wycoff and David Maister, *Owner Operator: Independent Trucker*, Lexington 1975.
American Trucking Associations, *American Trucking Trends*, Washington 1976.
______, *Annual Report*, Washington 1977.
Handling and Shipping, July 1976, *25*, 23–27.
U.S. Interstate Commerce Commission, "Empty/Loaded Truck Miles on Interstate Highways during 1976," Bureau of Economics, Washington, Apr. 1977.

Competition and Value of Service Pricing in the Trucking Industry: Reply

By A. S. De Vany and T. R. Saving*

Kenneth Boyer's comment is directed to our model of a competitive trucking industry operating in an uncertain market. In this model firms operated over a route $A-B$, carrying loads with value η_A from A to B (the fronthaul) and loads with value η_B from B to A (the backhaul). Under the assumptions that all loads at A and B had the common value η_A and η_B, respectively, and all loads were truckloads, we were able to show that, 1) the front- and backhaul would always be jointly served by the competitive firm, 2) that the firm's expected output is strictly less than its mean production capacity, 3) that expected wait for a load to be shipped is strictly positive, 4) that the competitive outcome is efficient, and 5) that competitive prices have the solution:

$$(1) \qquad p_A = C_\alpha + \alpha\eta_A \frac{\partial \overline{W}_A}{\partial \alpha} - \beta\eta_B \frac{\partial \overline{W}_B}{\partial \alpha}$$

$$(2) \qquad p_B = C_\beta + \beta\eta_B \frac{\partial \overline{W}_B}{\partial \beta}$$

where p_A, p_B = trip price at A, B; α, β = mean output from A to B, B to A; C_α, C_β = marginal hauling cost A to B, B to A; and $\partial \overline{W}_A/\partial\alpha$, $\partial \overline{W}_B/\partial\beta$ = marginal waiting time at A and B, respectively.

In interpreting the pricing equations we noted that price exceeds marginal hauling cost by an amount which we called a congestion toll. The congestion toll just equaled the costs imposed on all users of the system when the steady-state expected output increased at either A or B, capacity held constant. In the fronthaul market, those costs are the cost of increased delay at A consequent to an increase in α, less the benefits at B of receiving more trucks from A, which reduces delay in the backhaul market. For the backhaul market, the cost of a unit increase in the steady-state output rate is the increase in waiting cost at B alone. We further noted that competitive prices just equaled marginal social cost, which is marginal hauling cost plus marginal waiting cost. Since the latter is directly related to commodity value, we made the observation that competitive trucking prices reflect the value of the commodity shipped, a natural consequence of the fact that shippers bid prices up until the trucking industry is induced to hold just that level of capacity that minimizes the shipper's total cost of shipping; including price and inventory holding cost.

We made one more observation which is apparently controversial. We said value-of-service pricing, which we specifically defined to be pricing which is related to commodity value, is a natural consequence of the competitive process, and that one cannot say that the current pricing structure in trucking is discriminatory simply because it reflects commodity value. That is, we made the observation that value-of-service pricing is a necessary but not sufficient condition for the proof of noncompetitive behavior.

In response to this proposition, Boyer makes three claims. First, that our analysis is incorrect when the assumption that firms operate in only one market is dropped. Second, that observations on the common carriers provides evidence that service quality is independent of value of commodity. Third, that even if our model were a true model of common carrier pricing, it only explains a small portion of the existing premiums paid by high-valued commodities.

As we shall demonstrate below, Boyer's first point is simply wrong except in a very extreme case. His second point is based on

*Simon Fraser University and Texas A&M University, respectively.

evidence that the reader can evaluate as well as we so that we will not comment on this point. His third point is based on a hypothetical example that is very sensitive to the assumed quantity of shipments. Over and above any of these points, however, is the fact that we drew no conclusions in our paper concerning the efficiency of current pricing practices in trucking. Indeed, for us to have drawn the conclusion that "value-of-service pricing as currently practiced" is efficient would have been irresponsible since the paper was wholly theoretical and contained none of the detailed empirical work that would be required to render that judgment. On the other hand, to argue, as is often done in the trucking literature, that value-of-service pricing is "per se" discriminatory simply because it relates to commodity value is also unwarranted.

I. Single Market Firms

We can extend our single market per firm model by allowing each firm to ship all commodities. Let the expected demand for transport for each good depend on the full price of transporting that good:

$$P_i = p_i + \eta_i W_i \tag{3}$$

where P_i is full price, p_i is actual freight charge, η_i is holding cost, and W_i is the expected wait for the ith good.

On the supply side, assume that each firm orders the goods by priority (note current regulations on common carriers prohibit this type of system but our discussion pertains to an unregulated competitive system). Firms can select transport charges freely but must accept the consequence that the arrival rates of the various goods must be such that for each good the expected full price equals the market determined full price.

For this system, it can be shown that the optimal prices can be written as

$$p_i = C_\alpha + \sum_{k=i}^{n} \alpha_k \eta_k \frac{\partial W_k}{\partial \alpha_k} |\alpha_{ki}^*| \tag{4}$$

where C_α is marginal hauling cost, α_i is the expected arrival rate of good i, and $|\alpha_{ki}^*|$ is the absolute value of the total effect of an increase in commodity i shipments on the arrival rate of the kth commodity. From (4), it follows that competitive prices must exceed direct marginal hauling cost by a congestion charge. It also follows that the number of terms in the congestion charge decreases as we move to lower priority goods. Thus, congestion charges are greatest for the highest priority goods and then diminish.

It is easy to demonstrate that in a priority system the highest-holding-cost goods will have the highest priority. Then if depreciation and spoilage is ignored, priorities will be by value of truckload. But then hauling rates will be positively related to value of commodity shipped, that is, value-of-service pricing, even though risk pooling is allowed. The proof is simple and proceeds by contradiction. Let the firm be maximizing profits, and assume that all priorities are determined by holding cost, except that commodities i and $i+1$ are interchanged. Without loss of generality, we can consider adjacent commodities since any other pair can be represented by a series of adjacent transpositions. Consider the effect on profits of exchanging one truckload of good i with a truckload of good $i+1$. This is equivalent to changing the priority of a single load.

From the definition of full price and the fact that a single unit load from each category is being exchanged, we have that the change in profits ($\Delta\pi$) is

$$\Delta\pi = -(\eta_i \Delta W_i + \eta_{i+1} \Delta W_{i+1}) \tag{5}$$

But $\Delta W_i > 0$ since a good i load is being placed in priority $i+1$. In addition, $\Delta W_i = -\Delta W_{i+1}$ since good i and $i+1$ are exchanging expected transit times. Thus, we can write (5) as

$$\Delta\pi = (\eta_{i+1} - \eta_i)\Delta W_i \tag{6}$$

which is necessarily positive since $\eta_i > \eta_{i+1}$ by construction. Finally, if all priorities are by value of a unit load, then any exchange of unit loads between priorities results in (6) being negative since $\eta_i < \eta_{i+1}$ by assumption. Therefore, the profit-maximizing priority system must be by value of unit load.

Under what circumstances does the above system result in equal prices for all commodities? The answer here is that if the trucking firm has access to unlimited trucks, then no one waits and the congestion tolls all become zero. But it will never pay a trucking firm or a leasing firm to have so great a number of trucks such that buyers never wait. The reasons are that trucks are not free, and buyers have finite costs of waiting. In any case, Boyer's point does not depend on our assumption of single market firms but rather on our assumption of finite resources.

II. The Hypothetical Example

Boyer uses an ingenious technique to determine the congestion toll implied by our model. He ignores the fronthaul market and uses our price equation for the backhaul market (equation (2) above) to calculate hypothetical congestion tolls. By assuming values for β, η_β, and $\partial W/\partial\beta$ he calculates $\beta\eta_\beta(\partial W/\partial\beta)$ for an 800 mile route shipping 100 tons per day. His calculations are for different commodity values ranging from \$200/ton to \$5,000/ton. He concludes that the congestion tolls, which for marginal waiting time equal to unity range from 0.74 percent to 18.58 percent of the mean motor carrier charge per ton mile of 9.6 cents, cannot explain observed price differentials which may be as large as 400 percent.

In discussing these hypothetical examples, two points should be made. First, the congestion charges are not only sensitive to value, but to traffic flow and marginal waiting time. Second, traffic flow and marginal waiting time are endogenous to the system and cannot be independently chosen. Thus, the actual figures for the hypothetical example are meaningless. Accordingly, the results can be made to reflect any point of view with a proper choice of values for the parameters.

Consider Boyer's example wherein he assumes 100 tons shipped per day in trucks with capacity of 24 tons. These figures result in roughly 4 truckloads per day being shipped. Thus, he is considering a very small secondary market and our theory says the congestion tolls should be small if the primary market is large relative to the secondary market. On the other hand, let shipments be 10,000 tons so that truckloads per day are approximately 417 and we get a very different set of results for a marginal waiting time of unity as illustrated in Table 1.

As is easily seen from Table 1, the congestion tolls even for 1,000 tons shipped per day result in rates that will be some 300 percent higher for the highest-value commodity as compared to the lowest. Of course, these calculations prove nothing, but it is interesting that our model gives rise to rather substantial differences in commodity rates under reasonable assumptions.

TABLE 1—CONGESTION TOLLS FOR ALTERNATIVE VALUES AND TOTAL SHIPMENTS[a]

	Traffic Flow		
Commodity Value	100 Tons	1,000 Tons	10,000 Tons
\$200/ton	0.0713 (0.74)	0.713 (7.4)	7.13 (74)
\$500/ton	0.1783 (1.86)	1.783 (18.6)	17.83 (186)
\$1,000/ton	0.4458 (4.64)	4.458 (46.4)	44.58 (464)
\$2,000/ton	0.8917 (9.29)	8.917 (92.9)	89.17 (929)
\$5,000/ton	1.7833 (18.58)	17.833 (185.8)	178.33 (1858)

[a]The congestion charge as a percentage of the mean ton mile transport charge of 9.6 cents appears in parentheses below the congestion tolls.

Whether the theory explains the present rate structure in trucking is unknown. We would be surprised if it did because the common carrier obligation presently imposed causes the rate structure to diverge from a fully competitive system as noted above. Moreover, there may in fact be some discrimination in the present rate structure, though far less than has heretofore been supposed. The real test will come if trucking is eventually deregulated, for then it will more closely approximate the assumptions of our model. We are betting that deregulation will result in a proliferation of priorities of service, that there will be substantial variation of rates with a strong connection to commodity value, and that many shippers who have abandoned commercial trucking in favor of private fleets will come back.

III. Some Concluding Remarks

Boyer claims our model does not accord with what is observed in the regulated common carrier market. We agree, but it was never our intention to model this industry; we sought to model a competitive industry free of route restrictions and common carrier obligations. The industry that best comports with the model is the contract trucking industry where one commonly observes the specialization of firm to commodity which we assumed. Our argument is that where one sees widely varying commodity characteristics, one will see firms emerge which specialize in those commodities. Such a specialization essentially determines the fronthaul portion of the market, the backhaul will usually be a separate commodity. We see the assumption that each market has a single commodity as a practical first approximation that permits a useful analysis.

Regulated common carriers do not grant priorities to different commodity classes because they are prohibited from doing so. Such prohibition has been a major cause in the growth of the contract and own-account shipping markets. Where carriers are permitted to offer a range of service, one does see the phenomena we predict, namely high-value goods paying high prices and receiving quicker service. For example, under the agricultural exemption, trucking rates for agricultural commodities are not regulated. We predict that these unregulated rates will be related to time value and the evidence bears us out. It is known, for example, that cherries, a time-sensitive commodity, bear higher rates than apples, and more valuable slaughter cattle move at higher rates than do feeder cattle. In other areas as well, it is common for higher-value goods to move at higher rates. For example, where delivery service is offered on a first day, second day, etc., priority basis, higher-valued goods comprise the bulk of goods moving at the premium rate (see Brian Bayliss and Stanley Edwards, Table 8c, pp. 76, 84).

Boyer asserts that there is no evidence that queues or shipping delays exist in trucking. This claim largely testifies to the bias in his survey which asks truckers if they are giving good service. Surveys of shippers reveal that service is not as good as truckers indicate. We do not disagree with Boyer's finding that common carriers do not give priority to high-value goods since that is their required duty. We take as more serious the objection that there is no waiting for shipment, that is, there are no queues. Taken literally, Boyer is saying that no shipment ever waits. If that is true, how can it be that surveys of shippers consistently reveal that dependability of delivery and availability of equipment rank high among the determinants of shipper choice of transport mode? (See Douglas Woods and Thomas Domencich; Bayliss and Edwards.) According to Bayliss and Edwards: "the most important reasons for choice of mode was either speed of delivery to meet customers requirements or ready availability when required" (p. 50). If there were no waiting, why would it make any sense for firms to offer different priorities of service; for example, priority service in Great Britain and "Blue Label" service by United Parcel?

What we believe our model does is to point the way clearly to what information is

necessary to show that rates are discriminatory. It is not enough to show that rates differ by value of commodity by an amount which exceeds the actual marginal hauling cost of different commodities. Some correction must be made for service differentials. Our results also do *not* imply that deregulation is unnecessary because the current rate structure is optimal. Rather our results show that competitive trucking is a viable market organization; that it will serve small markets; that current allegations that pricing under competition would be volatile and ruinous are unfounded. It is only with a thorough understanding of competitive markets that we can fully analyze the impact of regulation.

REFERENCES

Brian T. Bayliss and Stanley L. Edwards, *Industrial Demand for Transport*, London 1970.

K. D. Boyer, "Queuing Analysis and Value of Service Pricing in the Trucking Industry," *Amer. Econ. Rev.*, Mar. 1980, *70*, 174–80.

A. S. De Vany and T. R. Saving, "Product Quality, Uncertainty, and Regulation: The Trucking Industry," *Amer. Econ. Rev.*, Sept. 1977, *67*, 583–94.

Douglas W. Woods and Thomas A. Domencich, *Competition Between Rail and Truck in Intercity Freight Transportation, U.S. Department of Transportation*, Charles River Associates, Inc., Cambridge, Mass. 1969.

The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Comment

By JOHN S. AKIN AND IRV GARFINKEL*

Historically, empirical estimates of the returns to years of education have consistently indicated that the percentage returns to whites exceed those to blacks. Possible explanations for the gap include discrimination and differing quality of education obtained during the years of attendance. In his 1973 article in this *Review*, Finis Welch suggested that differing quality of education is a major factor causing the observed gap. He provided evidence that education quality of young blacks and whites had been more equal than for older cohorts, and that young blacks who entered the labor force in the 1960's were experiencing a larger percentage return to an added year than were similar young whites. An implication of acceptance of Welch's hypothesis, combined with the reality of essentially equal returns for the "equally educated" young blacks, would be equal estimated returns to older cohorts in an empirical model correctly specified to control for socioeconomic backgrounds, experience, education years and education quality, Unfortunately Welch had no quality data for his sampled earners and therefore was unable to test directly for a major reduction in the gap for older cohorts that would be provided by correcting for the quality of their education.

Charles Link, Edward Ratledge, and Kenneth Lewis (hereafter, L-R-L) empirically tested a model in which quality is included by interacting it with years of education. Unfortunately they do not include either years or quality in a noninteracted form.[1] Most important, they use a sample containing 17–27-year-old males and are therefore unable to provide much insight into the effect of quality on older groups.

In this paper we use survey data for a group of men ages 30–55 to test whether an empirical model including quality of education information will predict a smaller gap in returns to years of education for blacks and whites than will a model with no controls for quality. Because of the age range of our sample we are also able to compare the differences in rates of return by race for different age groups. Our results suggest that quality of education is a significant determinant of earnings for both races, but that quality-controlled returns to years of schooling still differ considerably for blacks and whites. We find a sizable gap between young blacks and whites which remains more or less constant over age groups. Our best point estimates suggest that the gap may narrow slightly for older groups.

I. Data

Most of our data comes from the Survey Research Center's *Panel Study*. The data has several strong points as well as the inevitable weaknesses.[2] We do, however, have the almost essential data on quality of education, ability, motivation, and socioeconomic background for our sampled

*University of North Carolina, Chapel Hill, and University of Wisconsin-Madison, respectively. Helpful comments were provided by Orley Ashenfelter, Jim Heckman, Richard Layard, and other participants in the Workshop in Labour Economics, London School of Economics. We also wish to thank the managing editor and two anonymous referees for their comments and suggestions. All errors and opinions are ours alone.

[1]In an earlier note, Link and Ratledge (1975a) also use the Parnes data and include quality of education, both alone and in a form interacted with school years. In footnote 9 they state that, "No interactions were found between the quantity and quality of education" (p. 347). We have no explanation for this apparent inconsistency in their empirical results.

[2]Our data source is the Michigan *Panel Study of Income Dynamics* for the years 1968–72 along with added data from the 1930–60 Censuses of Population and Biennial Surveys of Education.

males.[3] That we have longitudinal data for a five-year period allows us to use an average (or "permanent") wage rate figure in order to remove the effects of temporarily high- or low-paying jobs for sampled individuals. Perhaps most important we have a sample of men chosen to represent the prime earnings years from age 30 to 55 (who attended school from the 1920's through the 1960's), and we have separate school expenditure data for black and white segregated school systems during the years in which certain southern states had such dual systems.[4] Men who are self-employed or who have zero earnings are excluded from the sample.[5]

In addition to the data obtained from the *Panel Study*, we obtained data from the 1930, 1940, 1950, and 1960 Biennial Surveys of Education on per pupil expenditures. From the 1930, 1940, and 1950, Biennial Surveys of Education, we also obtained per pupil school expenditures by race for the seventeen southern states with separate school systems prior to 1954. Values for these data during intercensal years were obtained by straight line interpolation. Each individual in the sample was then assigned the per pupil expenditure for the state in which he lived when he was 12-years old.

The measure of parent's income class in the *Panel Study* data is crude. Respondents were asked if their parents were poor, middle income, or upper income. Fifty percent said their parents were poor. Consequently, in addition to this measure we used the occupation of the respondent's father to create a more continuous and refined measure of parent's income. The income figure assigned was taken from the median earnings of the ten occupation categories as reported in the 1950 Census.

All monetary variables are deflated both over time and cross sectionally by cost-of-living deflators. The Bureau of Labor Statistics cost-of-living index was used to inflate all variables to 1972 dollar terms. A 1960 state cost-of-living deflator was used to deflate the state per pupil expenditure and father's income variables.[6]

II. Empirical Results

We have estimated the following three models for both blacks and whites:

Welch:

$$(1)\quad \ln WR = a_1 + b_1 S + c_1 ExP + d_1(ExP^2) + e_1(S\cdot ExP) + f_1 Z$$

Akin-Garfinkel:

$$\ln WR = a_2 + b_2 S + c_2(ExP) + d_2(ExP^2) + e_2(S\cdot ExP) + f_2 Z + g_2 Q + h_2(S\cdot Q)$$

Link, Ratledge, and Lewis:

$$\ln WR = a_3 + c_3(ExP) + d_3(ExP^2) + e_3(S\cdot ExP) + f_3 Z + h_3(S\cdot Q)$$

where WR = average wage rate over a five-year period

S = years of education

Q = education expenditure per pupil (quality)

$ExP = (\text{Age} - S - 5)$ = "experience"

Z = a vector of control variables measuring or indicating:

1) income of father of respondent
2) grew up in poor family
3) grew up in middle-income family
4) grew up in rich family
5) verbal ability (IQ) as determined from sentence completion test

[3]Our sample consists of 716 white male heads of households and 333 black male heads. These samples are weighted to sample sizes of 4,518 and 848, respectively, to make them nationally representative. We adjust the test statistics in order to prevent the expanded weighted sample sizes from affecting them.

[4]Our school expenditure data are state averages. We consider these data to be useful as school district averages. See Akin and Thomas Kniesner, and Link and Ratledge (1976).

[5]The self-employed are excluded because of the impossibility of separating the returns to labor from the returns to capital for this group.

[6]We are grateful to John Bishop for supplying us with this deflator. Interested readers may obtain details of Bishop's methodology from us.

6) achievement motivation index
7) grew up in farm family
8) years of education of father of respondent[7]

Our dependent variable is wage rate because we desire to measure the quality of the human capital possessed, not the intensity with which it is used.[8]

We also estimated the parameters of a more complete *ad hoc* model in which the interaction between quality and experience $(Q \cdot ExP)$ and the compound effect of quality, experience, and years of education $(S \cdot Q \cdot ExP)$ were added to Model 2. This model allows for all interactive and direct effects of quality on wage rates. The results from the estimation of this specification were very poor. None of the added interactions were statistically significant for blacks or whites, and the coefficient on school years became negative and insignificant for blacks and insignificant for whites. We therefore decided that the data were insufficient for estimation of the coefficients of this more complex model.

The results obtained from our three models are much as expected (see Table 1). The results from Model 2 indicate that when both Q and S are included along with their multiplicative interaction, the interaction is insignificant while each of the variables is highly significant with a positive sign. These results indicate that the coefficients on years of schooling are larger for both blacks and whites when quality of education proxies are added to the model. In Model 2 the returns to both quality and years of education are much smaller for blacks than for whites. The coefficients on the interactions terms provide evidence for the conclusion that marginal returns to years of schooling are reduced as both quality of schooling and experience increase.

Model 3 indicates that when the model contains no linear quality term the sign on the quality times years interaction becomes positive. This same result taken alone is interpreted by L-R-L as indicating that the returns to added years of education increase as quality increases. Given the results of Model 2 which show this interaction term to be insignificant when the two variables which comprise the interaction term are also included in the equation, this interpretation of the result seems questionable. We also note that even if the L-R-L model is accepted, our results, based on a sample which includes males of all prime working ages, indicate that the coefficient on the $(S \cdot Q)$ variable is considerably smaller for blacks than for whites.[9] The L-R-L results, based on a sample severely truncated in age (17–27), indicate that the coefficients are essentially the same. That our results are so different when we estimate essentially the same model but include men of all ages in our sample lends credence to our belief that it is impractical to test hypotheses about returns to the old as opposed to the young with a sample containing only the young. We also note that in our estimation of Model 3 the $(S \cdot ExP)$ coefficients are highly significant for both blacks and whites; again an opposite finding from their results.

Using these results it is a straightforward process to estimate returns to years of schooling in percentage terms. Using assumptions of eleven years of schooling and the natural logarithm of school expenditure (quality) equal to 5.0 (approximately the average for our total sample), in Table 2 we present estimated percentage returns for whites and blacks at ages 25, 40, and 55, and the white-black difference. The returns to education decrease with age for each racial group in each model, and with the exception of the Model 3 groups[10] the

[7] Our data and variables are described in detail in our earlier paper.

[8] See Alan Blinder for a discussion of the relative merits of wage rates and earnings as dependent variables. We ran similar regressions using earnings as the dependent variable and our results were not changed in any important ways from those reported in this paper. See our earlier paper.

[9] The difference between the two coefficients is significant at better than the .99 level on the basis of a one-tailed test, with the t-value computed as $(h_W - h_B)/(\alpha_W^2 - \alpha_B^2)^{1/2}$.

[10] Model 3 leads to the conclusion that there is essentially no gap by age 55 but a three-percentage point gap at age 25.

Table 1—Coefficients from Multiple Regression

Variable	Model 1: Welch W	Model 1: Welch B	Model 2: A-G W	Model 2: A-G B	Model 3: L-R-L W	Model 3: L-R-L B
S	.129[a]	.097[a]	.172[a]	.127[b]	–	–
ExP	.085[a]	.044[c]	.098[a]	.059[b]	.094[a]	.039[b]
ExP^2	−.0010[a]	−.0005	−.0010[a]	−.0005	−.0011[a]	−.0004
$S{\cdot}ExP$	−.0024[a]	−.0021[b]	−.0027[a]	−.0025[b]	−.0021[a]	−.0011[b]
Q	–	–	.361[b]	.172[a]	–	–
$S{\cdot}Q$	–	–	−.005	−.003	.023[a]	.015[a]
$Q{\cdot}ExP$	–	–	–	–	–	–
$S{\cdot}Q{\cdot}ExP$	–	–	–	–	–	–
R^2	.33	.29	.36	.33	.36	.32

Note: White Males (W); Black Males (B).
[a]Significant at the .99 level.
[b]Significant at the .95 level.
[c]Significant at the .90 level.

gap in percentage returns is always between 2 and 3.3 percentage points. The results strongly suggest that the gap in percentage returns to added education between whites and blacks narrows slightly with age. The results seem to hold in models with and without controls for school quality.[11]

To determine the exact relationship of the gap in returns to education associated with experience (or age) we only need examine the e_1 coefficients (on the ($S{\cdot}ExP$) variables). More formally stated, the difference between the e coefficient for whites and blacks is equal to the partial derivative with respect to experience of the white-black gap in percentage returns to years of education:

$$\frac{\partial\left(\dfrac{\partial \ln WR}{\partial S_W}-\dfrac{\partial \ln WR}{\partial S_B}\right)}{\partial \text{ Experience}}=e_W-e_B$$

For each of the three models the estimate of this slope of the gap between white and black earnings with respect to experience is[12] for Model 1, −.0003; Model 2, −.0002; Model 3, −.0010.

The signs and magnitudes of these derivatives suggests that the gap between black and white returns to years of education narrows slightly with age. Not only do Models 2 and 3, which include a quantity of education variable, exhibit this pattern of results, but so does Model 1, with no quality of education proxy. It appears that the use of a five-year average wage rate, combined with independent variables representing ability and motivation, with a sample of men over all prime working ages, leads to such results concerning the white-black gap irrespective of the model chosen. This narrowing of the gap with age in all three specifications is counter to expectations on the basis of the Welch "vintage" hypothesis. It should be noted again, however, that the

[11]It should be reiterated here that in each model we have controlled for verbal ability and achievement motivation. To the extent that these two characteristics are outcomes of added school years, our estimated returns to school years may be biased downwards. We choose to view these two variables as natural ability and personality proxies which are necessary controls for the estimation of returns to schooling. Given the size and significance of the estimated effect of school years on wage rates in each of our models, we doubt that the coefficients are biased downward to any large extent.

[12]Based on a one-tailed significance test, with the t-value computed as $t=(e_W-e_B)/(\gamma_W^2-\gamma_B^2)^{1/2}$, the difference between the e coefficients for blacks and whites is statistically significant in only the L-R-L model. This result suggests that the gap remains constant over time. Examination of realistic values for the e_i and the γ_i, however, leads us to believe that in our context this test is much too conservative. Either very large differences in rates of return or very small standard errors of coefficients are necessary for what would normally be a reasonable degree of significance. We therefore accept the coefficients as the best point estimates and use them as such in the earnings estimates that follow in this paper. That each e_i is highly significant within the context of its specific model adds to our willingness to accept these coefficients as best point estimates.

TABLE 2—ESTIMATED RATES OF RETURN= $\partial \ln WR/\partial S$

		Whites W	Blacks B	Difference $W-B$
Model 1:	Age 25	.107	.078	.029
	40	.071	.047	.024
	55	.035	.015	.020
Model 2:	Age 25	.123	.090	.033
	40	.082	.052	.030
	55	.042	.015	.027
Model 3:	Age 25	.096	.065	.031
	40	.065	.049	.016
	55	.033	.032	.001

amount of narrowing of the gap is very small in Models 1 and 2, and that the difference between the e_i coefficients is not statistically significant in any of the three models. (See footnote 11.) The results, however, remain counter to the Welch hypothesis even if we accept this insignificance and conclude that the gap remains the same for all age groups.

In order to express more graphically the differences in earnings represented by these relatively small percentage differences in returns, we estimate the earnings of blacks and whites at age 30 and 55, setting all non-age-related variables at their white sample mean values. We assume hours worked at the white sample mean also. These results indicate the earnings that would be predicted for both a black and a white of the given age if each had exactly the white sample average coefficients. All differences are, therefore, due to estimated parameter values. The results as presented in Table 3 are quite revealing.

The results in Table 3 emphasize a last point we wish to make. Even equal percentage returns to blacks and whites are not necessarily indicative of equal earnings for equally competent workers. The results of each of our models are disheartening when interpreted in absolute rather than percentage terms. At every age in every model, blacks with equal characteristics (including ability and motivation test scores!) are predicted to earn much less than whites. Even if quality being added to the model were to equalize *percentage* returns, as Welch seems to infer in his article, it would be difficult to

TABLE 3—BLACK AND WHITE PREDICTED EARNINGS WITH ALL CHARACTERISTICS EQUAL

Age	Model 1	Model 2	Model 3
30: W	7688	7818	7781
B	7245	7586	7689
55: W	9246	9257	9115
B	6507	7370	7755

rationalize that a 5 percent of $7,500 ($375) return to an added year of education for blacks at age 55 as opposed to a 5 percent of $9,000 ($450) return to whites with the same characteristics indicates the end of discrimination. That the comparisons are generally made in "percentage of base earnings" terms is unfortunate because of the misimpression that may be presented.

III. Conclusion

We have used survey data to estimate returns to added schooling for three distinct single equation models. Our estimated returns to education at age 40 are in the range of 7 percent of wage rate for whites and 5 percent of wage rate for blacks. These results are obtained from each model with controls for ability and motivation included. In two of the models we also control for quality of education.

Our major objective is to test the Welch "vintage hypothesis" which suggests that the gap between percentage returns to years of education for blacks and whites has narrowed for younger groups because of a relative upgrading of education quality for blacks. We find that quality has significantly affected wage rates for both races.[13] Even after controlling for quality, ability and motivation, however, we still find a large difference in predicted earnings for whites and blacks. We also find that the percentage returns to years of education for whites are approximately 3 percentage points greater at every working age. These

[13]In our earlier paper we showed that the rate of return to school quality is generally greater for blacks than for whites. The Model 3 results suggest that the gap narrows greatly with age and disappears by age 56.

results suggest that the gap is not narrowing for younger age groups, and that, in fact, it may be widening slightly for the young. Based on our best estimates for coefficients, the gap in fact narrows to a slight extent with age.

REFERENCES

J. Akin and I. Garfinkel, "School Expenditure and the Economic Returns to Schooling," *J. Hum. Resources*, Fall 1977, *12*, 460–81.

______ **and T. Kniesner,** "Proxies for Observations on Individuals Sampled from a Population," *J. Hum. Resources*, Summer 1976, *11*, 411–13.

A. Blinder, "On Dogmatism in Human Capital Theory," *J. Hum. Resources*, Winter 1976, *11*, 8–22.

C. Link and E. Ratledge, (1975a) "The Influence of the Quantity and Quality of Education on Black-White Earnings Differentials: Some New Evidence," *Rev. Econ. Statist.*, Aug. 1975, *57*, 346–50.

______ **and** ______, (1975b) "Social Returns to Quantity and Quality of Education: A Further Statement," *J. Hum. Resources*, Winter 1975, *10*, 78–89.

______ **and** ______, "Social Returns to Quantity and Quality of Education: A Further Statement: Reply," *J. Hum. Resources*, Summer 1976, *11*, 413–19.

______, ______, **and K. Lewis,** "Black-White Differences in Return to Schooling: Some New Evidence," *Amer. Econ. Rev.*, Mar. 1976, *66*, 221–23.

F. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *67*, 893–907.

Survey Research Center, *A Panel Study of Income Dynamics: Study Design, Procedures Available Data*, Ann Arbor 1972.

The Quality of Education and Cohort Variations in Black-White Earnings Differentials: Reply

By FINIS WELCH*

John Akin and Irv Garfinkel (hereafter, A-G) average a five-year panel data set into a single cross section and with it accomplish two things: First, if you take their preferred results literally, they dispel any notion that pupil expenditures on schooling approximate school quality. Second, they demonstrate what by now is a familiar story that cohort and life cycle effects are not easily disentangled in cross-sectional data.

Akin and Garfinkel measure quality of schooling by state average (regional and time price adjusted) per pupil expenditures on public elementary and secondary schools in an individual's state of residence at age 12. They present regressions first showing that pupil expenditures are significantly correlated with wage rates and then contrast a somewhat more flexible form of the wage-pupil expenditure relation with the form estimated and reported by Charles Link, Edward Ratledge, and Kenneth Lewis (hereafter, L-R-L).

Several points are relevant:

1. For whites, the A-G estimates of their flexible specification are not conceivably statistically different from the restricted L-R-L form which A-G criticize. For blacks, the specifications are probably not statistically different.[1]

2. The A-G "preferred" estimates imply a counterintuitive relation between income and quality of schooling. I presume that school quality refers to a rate of learning and that quality is worth more to those who get the most from it. The A-G estimates (their Table 1) contain statistically insignificant negative interaction between "quality" and years of schooling in which both quality and schooling have independent (significant) positive effects. If we assume that the statistically insignificant coefficient is zero, we have the anomalous result that an increase in quality is schooling neutral; that is, has an equiproportionate wage effect at all schooling levels. The point estimates with negative interaction are more perverse.

It is tempting to argue that state average pupil expenditures do not measure school quality and to suggest an alternative interpretation of the neutrality finding: that tax-based school expenditures are high in states where income is high. True, the A-G measure refers to residence at age 12 rather than current residence, but migration or switches between current and residence at age 12 are not reported. But—

3. Akin and Garfinkel's Table 2 contrasts schooling coefficients at various ages between Model 1 which excludes the expenditure variable and Model 2 which includes it. Inclusion of the expenditure variable increases the schooling coefficient! Since expenditures have an independent positive effect, the evidence is that schooling and the quality proxy are negatively correlated.[2]

*University of California-Los Angeles. I am indebted to James Smith for his suggestions.

[1]Akin and Garfinkel do not report statistics for the joint hypothesis implicit in contrasts of Model 2 (their preferred form) and Model 3 (the L-R-L form), but R^2s are reported. In principle, the R^2s imply the relevant F-statistics, but Akin-Garfinkel round their reported R^2s to two decimal points. For whites, the R^2 difference between Model 2 and 3 is 0.00 and even the maximum potential rounding error could not result in a calculated F (2,701) that would be significant at 5 percent. For blacks, the R^2 difference is 0.01 which with rounding errors could have been anything between 0.005 and 0.015. Even so, a difference of 0.0128 would produce a just significant F (2,318). Of the 0.01 range of possible rounding errors, only 22 percent could produce a marginally significant difference. I suppose you could say that the probability of no significant difference between Models 2 and 3 is 0.78 percent!

[2]This correlation is, of course, partial, holding other explanatory variables in the wage equation constant.

The expenditure variable is constant within states (at a point in time) and is deflated both for price level changes through time and for regional price level differentials. And the evidence is that one or both of two patterns exist: Schooling levels and deflated expenditures are negatively correlated regionally as would be true if, for example, deflated expenditures were higher in the South and/or deflated expenditures have fallen through time as schooling levels have increased.

Finally, A-G want to interpret their cross-sectional evidence as relevant both to life cycle and cohort phenomena. The associated identification problem is obvious and their attempt to reconcile it by restricting all cohort effects to pupil expenditures on schooling is novel, but maybe a little naive.

Each of the three models they estimate shows clearly that within schooling classes (or holding schooling constant) black-white wage ratios are higher for young men than for older men. If this is a life cycle effect, that is, if wages of blacks fall relative to whites as cohorts age, we should know it. But the optimistic alternative that relative wages do not fall over the life cycle, instead cross-sectional profiles monitor cohort progress, seems worth investigating and for such comparisons more than one cross section is preferred.

Twice in this *Review*, first in 1973 in the paper A-G note and again in 1977 in a paper coauthored with James Smith, I reported evidence for two cross sections in which black-white wage ratios did not decline as cohorts aged. In each case, I interpreted the evidence of higher black-white ratios for younger men as cohort convergence, as evidence of blacks "catching up."

Since then, Smith and I have contrasted black-white wage ratios revealed in the first eight Public Use Files of the March *Current Population Surveys*. Even though these are the largest income surveys available on an annual basis, narrowly defined age or experience-by-schooling-by-race cells contain so few observations that black-white wage ratios are often erratic. We therefore used regressions to smooth the data but calculated separate equations for blacks and whites, and allowed regression intercepts and schooling coefficients to vary year by year. Some of the other coefficients were constrained to equality across paired years whenever preliminary checks suggested no real differences. Finally, we used regression imputations for smoothed estimates of black-white weekly wage ratios for various labor market or, more precisely, school-leaving cohorts.

Table 1 gives our basic estimates for each year, 1967–74, for men with eight, twelve, and sixteen years of schooling for each of four cohorts: beginning work in 1952, 1957, 1962, and finally in 1966. In these data, we find no clear life cycle changes in black-white wage ratios during the eight-year period. There is a hint that earnings in the early career may grow more rapidly for black than for white college graduates, and a hint that the converse may be true for those with eight years of schooling. For high school graduates, there are no calendar year or life cycle trends and, overall, the simple rule that as a cohort enters the job market, the black-white wage ratio first established is then maintained, appears reasonable. Panel A of Table 2 includes the eight-year average wage ratio for each of these cohorts.

The lower ratios for high school graduates than for either those schooled for eight years or for college graduates, points to lower (proportionate) returns to grade school and to higher returns to college for blacks than for whites. The fact that wage ratios are higher for more recent cohorts demonstrates convergence and is elaborated in Panel B where first differences are shown.

Note the pattern among schooling levels. In every case, cohort convergence is more rapid for college than for those not completing high school. If the cohort interpretation of these and, for that matter, every set of data I have seen is correct, then the market is treating more recent black cohorts as more productive relative to whites. The positive association of cohort improvement with schooling level is clearly consistent with black-white convergence in quality of schooling. Even though the data show cohort convergence, it is obvious that cohort

TABLE 1—ESTIMATED BLACK-WHITE WAGE RATIOS BY YEARS OF SCHOOL COMPLETED, 1967–74

Year Cohort Entered Job Market	1967	1968	1969	1970	1971	1972	1973	1974
			A. Years of School Completed = 16					
1966	.834	.885	.846	.908	.892	.984	.879	.880
1962	.806	.858	.807	.864	.843	.912	.855	.853
1957	.779	.829	.774	.818	.798	.842	.824	.810
1952	.760	.804	.758	.781	.769	.791	.792	.758
			B. Years of School Completed = 12					
1966	.853	.779	.807	.768	.852	.865	.809	.765
1962	.817	.764	.772	.752	.809	.829	.799	.768
1957	.781	.747	.744	.738	.771	.799	.785	.763
1952	.755	.734	.731	.731	.746	.783	.769	.747
			C. Years of School Completed = 8					
1966	.892	.863	.914	.859	.913	.876	.798	.794
1962	.860	.850	.870	.837	.871	.844	.799	.807
1957	.830	.837	.833	.817	.834	.819	.799	.813
1952	.808	.826	.814	.804	.812	.808	.797	.807

Source: Smith and Welch, (1979, Table 15).

effects refer to more than quality of schooling. Everything affecting productivity including such background factors as parental schooling and unobservables like employer perceptions of productivity is involved. Yet, I for one seriously doubt that the case for school quality can be "proven." There simply are too many competing but observationally equivalent explanations. The problem would be easier if we knew how to measure school quality and if alternative hypotheses could be reduced to simple empirical constructs.

But, I do not know how quality of schooling can be measured. If nominal characteristics of schools are important, then there is support for the quality view. The federal government stopped publishing statistics of segregated schools in 1954 with the Supreme Court's desegregation decision, although a few southern states continued to publish such statistics until the early 1960's. These data show a clear pattern of convergence between black and white schools in such characteristics as pupil expenditures, teacher salaries, teacher credentials, pupil-teacher ratios, days attended each year, failing or retardation rates, etc. throughout the twentieth century. More recently, school integration has played a direct role which may be especially important for college students.

Just as quality of schooling is hard to measure, the competing explanation of environmental reductions in discrimination and of government-induced effects through antidiscrimination legislation or executive orders of the president are hard to approximate empirically. This is not the place to summarize this kind of research, but two points should be made. First, there have been several attempts to quantify effects of the legislation and very few of these studies assign a major role to governments in reducing black-white differentials; second, the data suggest that both cohort convergence and its positive association with schooling predates the modern antidiscriminatory legislation.

Attempts like those of A-G to explain black-white differences in effects of schooling on earnings are always interesting. Yet, the fact that "control for pupil expenditures" seems not to affect these differentials raises more questions about the interpretation of pupil expenditures as a proxy for school quality than it answers about determinants of school contributions to earnings. Finally, in a single cross section like the A-G five-year average, the one-to-one linear correspondence between cohort and work experience creates a fundamental identification problem. In principle, it can

TABLE 2—BLACK-WHITE WAGE RATIOS FOR COHORTS BEGINNING WORK, 1952–66, BY LEVEL OF SCHOOL COMPLETION

Years of Schooling	A. Annual Average Ratios, 1967–74: Cohort began work in 1966	1962	1957	1952
16	.889	.850	.809	.777
12	.812	.789	.766	.750
8	.863	.842	.823	.810
	B. Cohort Differentials			
	1966–62	1962–57	1957–42	
16	.039	.041	.032	
12	.023	.023	.016	
8	.021	.019	.013	

Source: See Table 1.

be broken via parameterization of cohort effects as A-G do by treating them as a function of school expenditures. In practice, the success of such efforts is open to question. My own experience from contrasting cross sections at different points in time suggests that essentially all of the black-white difference in cross-sectional profiles among cohort-experience groups is a cohort effect. If so, then the fact that this differential persists after the introduction of the expenditure variable simply suggests that expenditures do not tell the cohort story.

REFERENCES

J. Akin and I. Garfinkel, "The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 186–91.

C. Link, E. Ratledge, and K. Lewis, "Black-White Differences in Returns to Schooling: Some New Evidence," *Amer. Econ. Rev.*, Mar. 1976, *66*, 221–23.

J. Smith and F. Welch, "Black-White Male Wage Ratios, 1960–1970," *Amer. Econ. Rev.*, June 1977, *67*, 323–38.

______ **and** ______, "Race Differences in Income: A Survey and New Evidence," in Peter Mieszkowski and Mahlon Straszheim, eds., *Issues in Urban Economics*, Vol. II, Baltimore 1979.

F. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.*, Dec. 1973, *63*, 893–907.

The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Reply

By CHARLES LINK, EDWARD RATLEDGE, AND KENNETH LEWIS*

In an article in this *Review*, Finis Welch suggested that recently the percentage returns to education have risen faster for blacks than for whites. Welch's conclusion was based on cohorts of males from the 1960 Census and the 1967 *Survey of Economic Opportunity*. He observed that in models where only years of education, but no quality measures, were included, returns to education declined for both blacks and whites in older cohorts. The interesting result from Welch's point of view was that the percentage returns to black education fell off *appreciably faster* compared to whites as the cohort aged. The implication is that blacks have been gaining with respect to whites as far as returns to education are concerned. Welch observed that this gain to blacks is likely to be due to a recent rise in the quality of education for blacks relative to whites. From a policy point of view, this leads to the conclusion that, however limited its role may be, education still is a potential policy tool with which to reduce observed racial earnings differences.

The basic problem with Welch's results is that the observed decline in the returns to education for older blacks could have occurred for a number of reasons which could not be accounted for because of the limited nature of the data. That is, the more rapid decline in the returns to education for blacks (compared to whites) could be due to: differential life cycle effects (for example, credential effects, decline in discrimination); higher returns to the quality of education for blacks vs. whites; or similar returns to the quality of education for blacks and whites but greater increases in quality for blacks compared to whites—especially for the most recent cohorts.

In order to shed light on some of the above issues, we reestimated the Welch model and published the results in this *Review*. Our model, however, included a measure of the quality of education—the expenditure per student in 1968 where the respondent attended high school. The sample, males 17–27 in 1969, was from the "National Longitutional Survey of the Labor Force" (*NLSLF*) which is conducted at the Ohio State University. The importance of the 17–27 age group is that it is one of the cohorts in which blacks are likely to be receiving returns to education commensurate with whites. The importance of having a measure of quality is that it allows the estimation of the differential returns to quality for blacks and whites. Our basic result was that returns to quality are similar for young blacks and whites. The latter finding supports the Welch hypothesis that the reason for the rise in black vs. white returns to education is that the quality of black education has risen more rapidly vis-à-vis whites.

John Akin and Irv Garfinkel (hereafter, A-G) extend our analysis using a sample of male heads of household aged 30–55 in 1972 from the Michigan *Panel Study of Income Dynamics*. Using estimated wage equations which include the quality of education, they attempt to find evidence of Welch's observed differentials over a wider age group and to comment on the rate of returns to quality for different age-race cohorts. Their basic findings, however, are counter to those hypothesized by Welch and ourselves.

The outline of the present paper is as follows. In Section I, the conceptual model

*Associate professor of economics; director, urban policy research, College of Urban Affairs and Public Policy; and associate professor of economics, University of Delaware, respectively. We would like to thank Arthur Hoerl for the use of the ridge regression software which he has developed at the University of Delaware. Also, we acknowledge the helpful assistance of Jerry Russo. Finally, we wish to thank John Akin and Irv Garfinkel for providing their data files to us.

and basic issues are reiterated. Section II updates our earlier results with models more comparable to A-G than was the case in the earlier analysis that used the 1971 *NLSLF* data. Section III is a critical analysis of the A-G framework. We demonstrate, with the *Panel Study* data used by A-G, that Welch's and our results are upheld. Section IV contains a summary of the results.

I. Methodological Issues

The basic A-G model, estimated separately for blacks and whites, takes the form

$$(1)\quad \ln WR = a + bS + c\cdot Exp + d(Exp)^2 + e(S\cdot Exp) + f\cdot z + g\cdot \ln Q + h(S\cdot \ln Q)$$

where the notation is that of A-G. The percentage returns to years of schooling and school quality can be written respectively as

$$(2)\quad \frac{\partial \ln WR}{\partial S} = b + e\cdot Exp + h\cdot \ln Q$$

$$(3)\quad \frac{\partial \ln WR}{\partial \ln Q} = g + h\cdot S$$

Welch hypothesized that the quality of education was negatively related to one's vintage. Thus,

$$(4)\quad \ln Q = \eta(Exp) \qquad \eta' < 0$$

The total derivative of (2) with respect to *Exp* (i.e., a "total vintage effect") is

$$(5)\quad \frac{d\left(\frac{\partial \ln WR}{\partial S}\right)}{d\,Exp} = e + h\cdot \frac{d \ln Q}{d\,Exp} = e + h\eta'$$

where $$e = \frac{\partial\left(\frac{\partial \ln WR}{\partial S}\right)}{\partial\, Exp}$$

holding $\ln Q$ constant. In this form, the Welch vintage effect can be broken down into differential cohort, life cycle, or credential effects, $e_w - e_b$; differential cohort returns to an extra unit of quality, $h_w - h_b$; and differential changes in the quality of education as one ages, $\eta'_w - \eta'_b$.

Akin and Garfinkel examine $(e_w - e_b)$ for each of their three models to test Welch's basic contention that blacks for the past several years have finally begun to receive monetary benefits from education, in percentage terms, comparable with those for whites. They find estimates of $(e_w - e_b)$ to be negative or not significantly different from zero and interpret the results to be counter to Welch's fundamental observation. We would argue, however, that such a test should consider only the differences in *total*, rather than partial, vintage effects. In this sense, only the results of their first model, which they label "Welch" (a reduced-form equation for which $e_w - e_b$ measures total differential effects) are of interest. In that model $(e_w - e_b) = -.0003$, which is inconsistent with Welch's primary observation.[1] However, empirically, in A-G as well as in our paper, the total effects (their Model 1) and the partial effects (their Models 2 and 3) are similar.

Akin and Garfinkel also indicate that a major objective of their work is to test the Welch vintage hypothesis that the gap between black-white percentage returns to years of education has narrowed because of a relative upgrading of educational quality for blacks. We note, however, that A-G, by not distinguishing clearly between partial and total effects, cannot test the issue in which Welch was most interested, namely that the quality of black education has risen faster for blacks than whites in the most recent years.[2]

[1]Assuming perfect model specification, the *total* effect (i.e., equation (5)) can be estimated by looking at $e_w - e_b$ in the reduced-form equation which A-G label "Welch." That is, the partial and total derivatives of the returns to schooling with respect to *Exp* are identical in this reduced form. To infer the total derivative from models which include quality variables, one can set $e_w - e_b$ for the reduced form equal to the expression in (5).

[2]By comparing total vintage effects from the reduced form to partial effects from A-G's Models 2 and 3, one can make inferences about the upgrading of educational quality for blacks. To do so, however, would put extreme pressure on the model specification. It is possible that better data and estimation techniques in the future may allow more precision in separating these effects.

II. Updating our 1976 Article

We estimate four different wage equations, using data from the 1971 *NLSLF* for 1,611 whites and 516 blacks who were aged 19–29.[3] The first three models are essentially the ones A-G label "Welch," "Akin-Garfinkel," and "Link, Ratledge, and Lewis" (we label the last L-R-L:A, (see their Models 1–3). We also estimate a fourth model that we label L-R-L:B, which in the notation of A-G, takes the form

$$ln\, WR = a_4 + b_4 S + c_4(Exp) + d_4(Exp)^2 + e_4(S\cdot Exp) + f_4(Z) + h_4(S\cdot Q)$$

where $ln\, WR$ is the logarithm of the 1971 hourly wage rate. *Exp*, potential labor market experience, is defined as age minus education minus 5, subject to the constraint that only experience after age 16 is counted as useful experience. Quality Q is defined as the logarithm of the expenditure per student in 1968 in the school district where the individual received his high school education.[4] The Z term includes a variable representing years of schooling received by the father and a variable indicating one's IQ.[5] Table 1 presents the results of the estimations for each of the four models as well as estimates of the percentage returns to quality and years of schooling (i.e., $\partial\, ln\, WR/\partial S$ and $\partial\, ln\, WR/\partial\, ln\, Q$).

Three basic results are found in Table 1. First, the percentage returns to years of schooling of young blacks are comparable to young whites (see bottom of table). Second, the returns to quality, that is the natural logarithm of expenditures per pupil, when evaluated at the mean level of education (12.5 for whites; 10.8 for blacks) indicate higher returns to quality for blacks than whites. Third, contrary to A-G's findings, the estimated $e_w - e_b$ indicate, *ceteris paribus*, that the returns to education for blacks are less favorable in the older portion of the cohort. For each of the models, the estimate of this partial derivative is for Model 1, .0022; Model 2, .0024; Model 3, .0028; Model 4, .0026.

One final note on Table 1 involves the proper specification of the earnings relation. Conceptually, it seems that quality should enter interactively with years of schooling that is, Model 3 or Model 4. This specification allows the weighting of years of schooling by the quality. Obviously, expenditure per student is an imperfect measure of school quality but by interacting S and Q, one can account for the number of years that a student receives that particular level of expenditure. The additive term does not allow for this.[6] That is, the only reason quality should be in the equation is to weight the years variable to attempt to make a year in one school comparable to that in another. We fail to see how the additive term accounts for such a weighting.

A potential problem that arises in such an interactive framework, i.e., $S\cdot ln\, Q$, is that by multiplying post-high-school years into the $ln\, Q$ variable, we are probably overstating the impact of the quality of elementary and secondary education on earnings. To account for this, the regressions, not reported, for Models 2–4 were rerun where years of education in the interaction term is limited to twelve. The returns to education of young blacks are still comparable to those of their white counterparts. Also as expected, the returns to quality drop, especially in Models 3 and 4, with the drop offest by a rise in the coefficient of the years of education variable.

[3]Included in our sample are males who worked at least thirty weeks in the previous year, usually worked at least thirty hours per week, had a positive wage rate, and were not enrolled in school. In order to account for missing data in the IQ, education of father, and expenditure variables, means were inserted. Dummy variables created to test for bias as a result of inserting means were statistically insignificant and dropped from the equation.

[4]See our paper, p. 227, for comments on the limitations of this measure.

[5]See R. Herriott and A. Kohen for a discussion of this IQ measure.

[6]Even if the school quality variable was measured correctly, since the measure attributed to the student is for only one year (a problem also with A-G's measure) unless the student does not migrate, the quality measure still will be subject to sizable measurement error.

TABLE 1—REGRESSION RESULTS FOR MALES AGE 19–29: *NLSLF* DATA
(t-statistics in parentheses)

Variable	Model 1: Welch *W*	*B*	Model 2: A-G *W*	*B*	Model 3: L-R-L:A *W*	*B*	Model 4: L-R-L:B *W*	*B*
S	.0654	.0826	.2874	.2948	–	–	−.0007	−.0125
	(7.04)	(5.00)	(3.26)	(1.06)			(.04)	(−.33)
Exp	.0920	.0693	.0911	.0683	.0897	.0750	.0902	.0709
	(3.48)	(1.70)	(3.47)	(1.69)	(3.82)	(1.95)	(3.43)	(1.75)
$(Exp)^2$	−.0034	−.0017	−.0034	−.0017	−.0034	−.0018	−.0034	−.0017
	(−3.53)	(−1.01)	(−3.60)	(−.97)	(−3.65)	(−1.07)	(−3.54)	(−.99)
$S \cdot Exp$	.0001	−.0021	.0002	−.0022	.0002	−.0026	.0002	−.0024
	(.05)	(−1.05)	(.12)	(−1.06)	(.18)	(−1.36)	(.14)	(−1.16)
$\ln Q$	–	–	.6090	.5943	–	–	–	–
			(3.32)	(1.12)				
$S \cdot \ln Q$	–	–	−.0350	−.0339	.0103	.0142	.0102	.0159
			(−2.53)	(−.76)	(8.35)	(5.72)	(4.43)	(2.74)
$\bar{R}^2$	.24	.20	.25	.21	.25	.21	.25	.21
$\frac{\partial \ln WR}{\partial S}$	.066	.068	.067	.069	.066	.070	.066	.070
$\frac{\partial \ln WR}{\partial \ln Q}$	–	–	.1715	.2282	.1287	.1534	.1275	.1712

III. Reanalysis of the A-G Data

In this section, we reestimate the A-G models—with their data—in order to test the sensitivity of their results to some basic alterations in their variables, sample, weighting procedure, and pooling procedures.

Several comments on A-G's empirical strategy are in order. First, because of the nature of the *Panel Study* sampling procedure, it is necessary to weight the data so that the sample is representative of the *U.S.* population. Akin and Garfinkel chose a weighting procedure which results in the overweighting of the low-income population. Our weighting procedure weights the population so that each observation receives its proper weight with respect to the other observations in the sample.[7] As it turns out, the differences between the two weighting procedures cause only minor differences in the results.

A more important problem involves A-G's choice of dependent variables, that is, the five-year-average hourly wage rate. The procedure used to calculate the wage rate for an individual in a particular year included two steps. They first checked a variable which reported the respondent's regular hourly wage rate. If the respondent did not have a regular wage rate or did not satisfy their acceptance criteria, the hourly wage was calculated by dividing the head's annual labor earnings by the total number of hours. There are two problems with this procedure. First, a sizable proportion of the respondents did not answer the question about their regular hourly wage rate. Nonrespondents to this question appear to be whites. Secondly, the hourly wage rate reported in the *Panel Study* refers to the 1972 wage. If the annual earnings divided by total hours calculation is required, that wage refers to 1971 because annual earnings and total hours as contained in the 1972 record refer to 1971 annual earnings and hours. The end result is that A-G are mixing wage concepts and time periods in calculating their five-year-average wage rates for individuals in the sample. In contrast, the five-year-average hourly wage rate we use below is calculated by taking annual labor earnings in the preceding year and dividing this by total hours in the preceding year. In addition, heads of households making less

[7]The A-G procedure involved using the square root of the weight. Our weighting procedure takes the weight for each observation and divides it by the mean weight for that group so that the sum of the weights *equals* one and the relevant statistics are based on the actual number of cases. We would like to thank James Morgan for clarifying this issue.

TABLE 2—REGRESSION RESULTS FOR MALES (ALL AGE GROUPS): *Panel Study* DATA
(t-statistics in parentheses)

Variable	Model 1: Welch W	B	Model 2: A-G W	B	Model 3: L-R-L:A W	B	Model 4: L-R-L:B W	B
S	.0974	.1164	.0781	.2543	–	–	−.0470	.0547
	(5.25)	(3.22)	(0.90)	(3.62)			(1.55)	(1.45)
$S\cdot Exp$	−.0012	−.0025	−.0013	−.0048	−.0015	−.0010	−.0009	−.0025
	(1.83)	(2.11)	(1.85)	(3.62)	(2.99)	(1.98)	(1.30)	(2.19)
$ln\,Q$	–	–	.2607	.2874	–	–	–	–
			(1.54)	(3.36)				
$S\cdot ln\,Q$	–	–	.0047	−.0137	.0194	.0150	.0246	.0132
			(0.35)	(1.59)	(7.82)	(5.11)	(5.90)	(4.16)
$\bar{R}^2$	.29	.35	.32	.40	.32	.38	.32	.38
n	700	305	700	305	700	305	700	305

than \$.50 per hour were eliminated from the sample. These extreme values were left in the A-G data set.

The final data problem concerned A-G's formulation of the years of education variable. They assigned zeros to the missing data codes and used what we thought were inconsistent codes for college "B.A. with no advanced degree" and "graduate work with a professional degree." We eliminated observations that 1) had average wage rates of less than \$.50, or 2) had missing data for years of education, in order to derive our final sample of 700 white and 305 black heads of households.[8]

The A-G models (as well as L-R-L:B) were reestimated for a sample pooled over all age groups. Regressions are also reported for the following age groups: 30–35, 36–40, and 41–55. The results for the all-age-group regressions are shown in Table 2. Comparing the results of these regressions with those from A-G's Table 1, it becomes apparent that ours are more favorable to the Welch vintage hypothesis. The estimated $e_w - e_b$ are

	L-R-L Table 2	A-G Table 1
Model 1	+.0013	−.0003
Model 2	+.0035	−.0002
Model 3	−.0005	−.0010
Model 4	+.0016	–

Model 3 provides the only contradiction to the observation that the returns to years of schooling become less favorable for blacks in older vintages, while the A-G model (Model 2) provides the strongest support for the vintage effect in the all-age-group regressions.

Even though the Table 2 coefficients are generally supportive of the observation that the returns to black years of schooling are less than those of whites, these results should be interpreted with caution.[9] We break the sample into three age groups: 30–35; 36–40; 41 and above. Conventional Chow tests for homogeneity in equation structure for all models indicate at the .01 level of significance that for blacks the age groups *cannot* be pooled. The tests for

[8]We coded these college graduate categories 16 and 18 instead of 15 and 17. The exclusions eliminated 16 whites and 28 blacks.

[9]In reviewing results from Table 1-5, the reader will notice that the A-G model tends to have insignificant coefficients on the key variables when compared to the other models. This is the result of an ill-conditioned matrix when all variables and their interactions are present. This is also the reason that A-G found large changes in the relative size of their coefficients upon introducing quality additively as well as interactively. In preparing this note, we estimated a parallel set of models using transformations on all interaction terms to insure that they were orthogonal to the component variables. The additive coefficients retain the statistical significance they had prior to the introduction of the interactive terms. The results of these orthogonal regressions are similar to those reported in our tables. We are indebted to Arthur Hoerl for his advice on this matter, and in particular for the use of his ridge regression software which in the process of estimating regressions by the ridge technique, orthogonalizes cross products and polynomials.

TABLE 3—REGRESSION RESULTS, *Panel Study* DATA, MALES
(t-statistics in parentheses)

Variable	Model 1: Welch W	Model 1: Welch B	Model 2: A-G W	Model 2: A-G B	Model 3: L-R-L:A W	Model 3: L-R-L:A B	Model 4: L-R-L-B W	Model 4: L-R-L-B B
Age 30–35								
S	.0817	.1192	.0332	−.0366	–	–	−.1664	−.0460
	(4.68)	(5.13)	(.14)	(.20)			(3.36)	(−1.30)
$ln\,Q$	–	–	.4551	.0192	–	–	–	–
			(.79)	(.05)				
$S\cdot ln\,Q$	–	–	.0113	.0287	.0180	.0247	−.0449	.0304
			(.26)	(.92)	(6.31)	(8.05)	(5.29)	(5.69)
$\bar{R}^2$	.21	.59	.33	.69	.28	.69	.33	.69
n	167	83	167	83	167	83	167	83
Age 36–40								
S	.0822	.0589	−.0503	.1592	–	–	−.0349	.0595
	(4.06)	(2.17)	(.21)	(1.32)			(.57)	(1.19)
$ln\,Q$	–	–	−.0347	.2185	–	–	–	–
			(.07)	(.91)				
$S\cdot ln\,Q$	–	–	.0239	−.0206	.0156	.0082	.0212	−.0001
			(.59)	(.85)	(4.60)	(1.78)	(2.01)	(.00)
$\bar{R}^2$	.39	.50	.40	.49	.41	.47	.41	.49
n	124	71	124	71	124	71	124	71
Age 41 and Up								
S	.0601	.0095	.0554	.1211	–	–	−.0462	−.0073
	(6.96)	(.67)	(.61)	(2.27)			(−1.49)	(.25)
$ln\,Q$	–	–	.2405	.2910	–	–	–	–
			(1.20)	(2.82)				
$S\cdot ln\,Q$	–	–	.0013	−.0233	.0118	.0024	.0197	−0036
			(.08)	(2.14)	(7.77)	(.92)	(3.56)	(.67)
$\bar{R}^2$	.27	.26	.29	.29	.28	.26	.29	.25
n	409	151	409	151	409	151	409	151

whites were not as dramatic but the differences were significant at the .01 level. Table 3 presents the relevant regression coefficients for the three respective age groups.

The only cohort difference interaction A-G utilizes is $S\cdot Exp$, which constrains every other coefficient to be the same for all age groups. Analyzing the results by age group, which was the Welch approach in his original article, allows us to examine the pattern of returns for each of the cohorts when the coefficients of all variables are not constrained to be the same for different age groups. In essence, this allows the specification for $\partial\, ln\, WR/\partial S$ and $\partial\, ln\, WR/\partial\, ln\, Q$ to vary in a non-linear fashion between cohorts. Because we truncate the sample in fairly narrow age groups, the $S\cdot Exp$ term which appears in the all-age regressions is deleted from the model.

The basic results from the *NLSLF* and the *Panel Study* data are brought together in Table 4, where pooled-age-group and by-age-group calculations are presented for the percentage returns to schooling and quality for each of the four models. These calculations are based on the mean values for $ln\, Q$, Exp, and S for the cohort in question. The patterns of the returns are suggestive. Irrespective of the model, the percentage returns to years of schooling are more favorable to blacks in the 19–29 and 30–35 age cohorts with the difference becoming more and more unfavorable to blacks in the two oldest age cohorts. In fact, the coefficient of S in Model 1 for the age 41 and above cohort is significantly lower for blacks. Similarly, the percentage returns to quality favor blacks in the 19–29 cohort and whites in the older age groups. The one exception is for Model 3 in the 30–35 age

TABLE 4—RETURNS TO SCHOOLING AND QUALITY

	Model 1: Welch			Model 2:A-G			Model 3: L-R-L:A			Model 4:L-R-L:B		
	W	*B*	Difference	*W*	*B*	Difference	*W*	*B*	Difference	*W*	*B*	Difference
Percent Returns to Schooling ($\partial lB\ WR/\partial S$)												
Pooled	.067	.053	.014	.081	.062	.019	.072	.053	.019	.071	.059	.012
19–29[a]	.066	.068	−.002	.067	.069	−.002	.066	.070	−.004	.066	.070	−.004
30–35	.082	.119	−.037	.101	.124	−.023	.108	.138	−.030	.102	.124	−.022
36–40	.082	.059	.023	.089	.049	.040	.091	.044	.047	.089	.059	.030
41+	.060	.010	.050	.063	.011	.052	.065	.011	.054	.063	.010	.053
Percent Returns to Quality ($\partial\, ln\ WR/\partial\, ln\ Q$)												
Pooled	–	–	–	.319	.150	.169	.239	.150	.089	.303	.132	.171
19–29[a]	–	–	–	.175	.228	−.053	.129	.153	−.024	.128	.171	−.043
30–35	–	–	–	.601	.345	.256	.232	.280	−.048	.578	.345	.233
36–40	–	–	–	.266	.019	.247	.196	.080	.116	.267	−.001	.268
41+	–	–	–	.256	.077	.179	.141	.022	.119	.236	.033	.203

[a]The 19–29 age group results refer to the "young men" from the *NLSLF*.

group. Black percentage returns to quality are significantly higher than those of whites in the 19–29 and 30–35 age groups and significantly lower in the age 41 and above category.[10] A word of caution is in order when evaluating the percentage returns to schooling and quality at their means, since the percentage returns to blacks in the two youngest age groups would look even more favorable if common means were used in the calculations.[11] In summary, the results are consistent with the idea of the Welch vintage effect. That is, the returns to schooling and the quality of schooling are most favorable to blacks in the two youngest cohorts.

Akin and Garfinkel make the very important point that even comparable percentage returns may not necessarily indicate equal wage rates for equally competent workers. Thus they recast their results in absolute terms by computing the implied average wage rates (earnings) for whites and for blacks aged 30 and 55, setting all non-age-related variables at their white sample means in the respective wage equations. As Stanley Masters has noted elsewhere, however, this approach involves the dubious assumption that whites could be given the black earning function without changing that function. In Table 5 we compute the implied black-white wage rates for various ages by setting all non-age-related variables at their *black* sample means. The computations are based on the by-age-groups models using the *Panel Study* data. Akin and Garfinkel find that whites aged 30 and 35 in 1972 with equal characteristics are predicted to have higher wage rates than blacks. Our results are not so clear-cut. For males aged 32 in 1972, blacks are predicted to earn higher wage rates than whites with the same characteristics for two of the four models. Comparing the differences across age groups, the implied wage gap has widened in favor of whites in older cohorts, a result similar to A-G, but consistent with Welch's vintage hypothesis.

IV. Summary

In this paper we have examined two data sets in order to further examine Welch's contention that blacks for the past several

[10]The significance levels for the differences in the 19–29, 30–35, and 41 and above age groups are .07, .07, .01, using a one-tailed test and standard covariance procedures. The results when the alternative $S\cdot ln\,Q$ formulation is employed (i.e., S can take a maximum value of 12) yield the same basic conclusions as those indicated in Table 4.

[11]The mean levels of education for blacks and whites in the age 30–35 age group are 11.34 and 12.88, respectively. Actually, the racial patterns of the means of S by age groups are suggestive of why pooling is inappropriate: 36–40, whites 12.6 and blacks 9.7; age 41–55, whites 12 and blacks 9.1.

TABLE 5—IMPLIED WHITE AND BLACK HOURLY WAGE RATES WITH ALL CHARACTERISTICS EQUAL: *Panel Study*, BASED ON BY-AGE-GROUPS EQUATIONS (TABLE 3)[a]

	Model 1: Welch			Model 2: A-G			Model 3: L-R-L:A			Model 4: L-R-L:B		
	W	*B*	Difference	*W*	*B*	Difference	*W*	*B*	Difference	*W*	*B*	Difference
Age 32	3.80	3.30	.50	3.10	3.30	−.20	3.60	3.30	.30	3.20	3.30	−.10
Age 37	3.50	3.00	.50	3.20	3.00	.20	3.20	2.90	.30	3.20	3.00	.20
Age 47	4.00	3.20	.80	3.30	3.20	.10	3.70	3.20	.50	3.60	3.20	.40

[a]Implied hourly wage rates for whites are based on the insertion of non-age-related black means in the white wage equation. Implied wage rates for blacks are based on the insertion of black means in the black equation.

years have begun to receive monetary benefits from education commensurate with those of whites and, secondly, that the gap has narrowed because of a relative upgrading of educational quality for blacks.

Counter to A-G and in support of Welch, we find that the gap in percentage returns to schooling has narrowed for blacks in younger cohorts. Second, contrary to A-G and in support of Welch, we find percentage returns to schooling and to quality for the younger black cohorts to be commensurate with those for whites. In fact, we find percentage returns to schooling and to quality to be somewhat more favorable for young blacks than for whites.

Akin and Garfinkel are quick to point out, however, that even if *percentage* returns are comparable, absolute wage rate differences may still favor whites. Thus they warn that equal percentage returns to blacks and whites are not necessarily an indication of equal earnings for equally competent workers. While A-G compute implied wages to be greater for young whites than blacks, our results are not as disheartening and are mixed, depending upon which model specification is adopted. Consistent with Welch, however, we find a narrowing in the implied black-white hourly wage gap for younger cohorts.

REFERENCES

J. Akin and I. Garfinkel, "The Quality of Education and Cohort Variation in Black-White Earnings Differentials: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 186–91.

R. Herriott and A. Kohen, "On Pooling of Mental Ability Measures for Different Tests: A Pragmatic Approach" unpublished paper, Ohio State Univ. 1973.

A. Kohen, "Determinants of Early Labor Market Success Among Young Men: Race, Ability, Quantity and Quality of Schooling," unpublished doctoral dissertation, Ohio State Univ. 1973.

C. Link, E. Ratledge, and K. Lewis, "Black-White Differences in Returns to Schooling: Some New Evidence," *Amer. Econ. Rev.*, Mar. 1976, *66*, 221–23.

Stanley Masters, *Black-White Income Differentials*, New York 1975.

F. Welch, "Black-White Differences in Returns to Schooling, *Amer. Econ. Rev.*, Dec. 1973, *63*, 893–907.

Center for Human Resources Research, "National Longitudinal Survey of the Labor Force," Ohio State Univ. forthcoming.

Survey Research Center, *A Panel Study of Income Dynamics: Study Design, Procedures Available Data*, Ann Arbor 1972.

Measurement of Tax Progressivity: Comment

By DAVID G. DAVIES*

In a recent issue of this *Review* Daniel Suits attempts admirably to attack the old problem of discovering a single figure, statistic, or index to describe whether a tax (or a tax system) is progressive, proportional, or regressive.[1] His ingenious index is inspired by and related to the Gini ratio. The gist of Suits' contribution is best described by reproducing and using his Figure 2 (shown herein as Figure 1) as well as his own words.

Suits notes the following properties about this figure: "Analogously to the Gini ratio we define the index of progressivity S in terms of K, the area of the triangle OAB, and L, the area $OABC$, contained between the Lorenz curve and the horizontal axis OA, so that

(1) $S=(K-L)/K=1-(L/K)$" (p. 749)

Suits goes on to explain that under a proportional tax, $L=K$, so $S=0$; but with a progressive tax the area L is smaller than K and $S>0$. With a regressive tax, $L>K$ and S is negative since the Lorenz curve lies above the diagonal.

I have several criticisms of the S index. The first is that Suits uses the annual family income of Joseph Pechman and Benjamin Okner as his proxy measure of economic power or economic well-being. Inspection of Figure 1 and equation (1) reveals that S is calculated by integrating over annual family income. This procedure introduces a bias into the magnitude of S and vitiates Suits' goal of providing a "...widely useful index of tax progressivity..." (p. 752).

As noted in my 1976 book, the lowest income class in any cross-sectional budget study based on the data of any one given year almost always has an average propensity to consume that is greater than one; often it is greater than two. When a family can spend twice its annual measured income on consumption, the concept of annual income is not a very useful indicator of the economic well-being of that family.

For individuals and households currently to consume more than their current income, the relationship between their past income and consumption must have been such that positive saving enabled the accumulation of assets which can now be partially or entirely liquidated; alternatively, there must be an expectation that future income will exceed consumption (so that liabilities may be incurred currently). As implied by the Modigliani-Brumberg-Ando "life cycle" hypothesis, the decision to reduce assets or incur liabilities stems from an attempt to stabilize one's consumption over the lifetime of a household.

Annual income figures tend to understate the long-run incomes of current low-income earners, while at the same time they tend to overstate the lifetime incomes of those individuals who currently receive high incomes. This proposition holds because current low-income recipients consist in a large measure of very young individuals and older retired persons.

The vast majority of retired and young individuals have not spent and will not spend most of their lives in the lower-income classifications. Joseph Califano, Secretary of Health, Education, and Welfare, has reported, for example, that "... over a six-year period (1967–1972), only 3 percent of the American population was poor in every one of those six years" (p. 5).

Reliance on annual figures overstates the average rate of tax that current lower-

*Professor of economics, Duke University. I would like to record my appreciation to my colleagues Bruce Bolnick, Martin Bronfenbrenner, and John Weymark, and especially to an anonymous referee.

[1]For examples of some earlier attempts to cope with this problem, see my 1961, 1971 papers, Richard Musgrave and Tun Thin, Musgrave, and Jeffrey Schaefer.

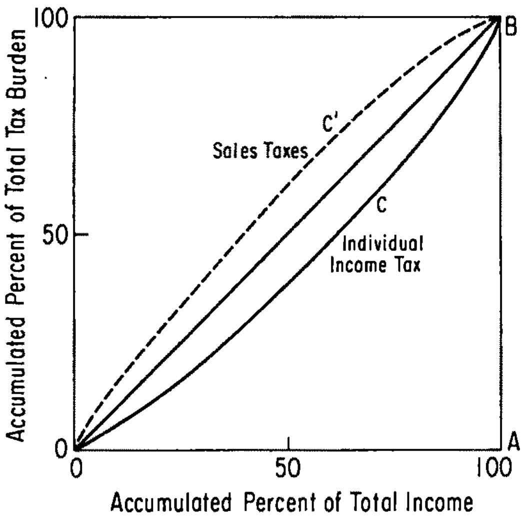


FIGURE 1

income recipients pay on their lifetime incomes, and understates the rate of tax of the higher-income earners. This conclusion follows from the fact that Suits derives his rate of taxation by dividing the amount of tax paid by the family's annual income. As a result the incidence of tax or tax structure is made to appear more regressive or less progressive than it actually is.[2]

Incidence and equity can best be measured by comparing different individuals' ratios of lifetime taxes paid to lifetime income earned, but this information is not available. Probably the next best practical solution to this problem is to use income concepts that incorporate a longer time horizon than one year. Irving Fisher's concept of income and Milton Friedman's idea of permanent income have been used with some success,[3] but the need for a longer time horizon in incidence studies is only just becoming generally recognized in the literature.[4]

[2]For a more extensive discussion of this point, see my 1963 paper.

[3]See my 1959 and 1960 papers, and Thomas Mayer for an analysis with several alternative concepts of income. The point is to utilize some notion of long-run income as a measure of economic well-being and have it serve as a basis for measuring progressivity, not to use it as a tax base to calculate actual taxes owed to a government.

[4]See, for example, Schaefer, Mayer, Pechman and Okner (pp. 8–9, 52, 67), and Henry J. Aaron (p. 92).

My second criticism of Suits' contribution is also concerned with the appropriate definition of income. It will be recalled that he uses Pechman and Okner's excellent data bank. As they point out, their definition of family income encompasses various receipts including Social Security benefits and cash welfare benefits, but it does not include the various and considerable in-kind payments that are now distributed primarily to low-income groups.[5]

John Korbel of the Congressional Budget Office has discovered that government cash assistance expenditures have grown by about four times during the decade following the implementation of the Great Society programs, but during this same period 1965–75, in-kind transfers increased by approximately 1,600 percent. Total expenditures for 1976 amounted to \$41.2 billion. He notes that the programs such as food stamps, medical care, and housing assistance are targeted primarily for the poor, but the data are "...not reflected in the official poverty statistics because these benefits are not counted as income" (p. xiv).

In order to obtain a more accurate value for S, this \$41 billion needs to be allocated properly and added to the income of the poor. This change will decrease the average rate of taxation for those individuals who had been classified into lower-income groups on the basis of money receipts only. Not accounting for in-kind transfers understates both the total income and magnitude of the S index, while counting them moves all tax lines toward point A or more progressivity on Suits' Figure 2 (my Figure 1).

If the only problems with the index S were using an income concept with an unreasonably short time horizon, and the exclusion of the large and increasing sums of in-kind transfers in family income, we could patch up the index simply by employing a longer-term concept of income and by including, as Edgar Browning does, the value of food stamps, Medicaid, and so forth into

[5]The theoretically appropriate "income base" is not yet settled. For a further discussion of the question see W. Irwin Gillespie, and Morgan Reynolds and Eugene Smolensky.

the reported family income recorded in budget studies.

A third problem with S, however, is inherent and not subject to equally easy correction. Suits notes that "The index S measures the average progressivity of a tax or tax system across the entire income range, yet some taxes are progressive over one range of incomes and regressive over another..." (p. 752). His Figure 5, which is reproduced here as Figure 2, compares two taxes one of which is progressive on higher incomes but regressive over the lower range, and vice versa. In both cases S takes on the value of 0 and Suits concludes that both taxes are proportional.[6]

This conclusion is at best a caricature of the real situation, because for each tax the only place over the entire range of incomes where there is proportionality is the point at which it crosses the diagonal. What grave errors a policymaker could make if we told him both tax A (dotted line in Figure 2) and tax B (solid curved line) were proportional when in fact A is progressive over the lowest half of society's income and regressive over the top half, the opposite holding for B. Would he not conclude in his mind that both A and B were described by the line OB which is a true proportional tax? Employing the S index in situations like these creates misinformation. Widely different configurations can yield the same value for S, and thus hide important knowledge and mislead us in our evaluation of how groups of families are faring. To be useful the data must be disaggregated so that the policymaker knows what is happening by income class or percentile of income or population. Using the S statistic in this situation is akin to employing an index to try to determine which of two intersecting demand or indifference curves is the higher.

Suits recognizes this problem when he states: "Distributions with equal means can have widely different variances; those with equal variances can have widely different skewness" (p. 752). Yet he does not appear to appreciate the many implications and limitations nor the narrowness of the applicability of the S index when trying to apply it to empirical taxes and tax structures. Real world taxes often do not act in a well-behaved manner. As Lester Thurow observed in his excellent review of the Pechman-Okner book, the variance of tax rates "...within each income class is much larger than the variance across income classes. The tax system is both progressive and regressive" (p. 191). Moreover, taxes frequently manifest a U-shaped tax-income relationship in which the tax is regressive at the lowest end, progressive at the highest end, and proportional or mildly progressive in the middle income range. P. Bently, D. J. Collins, and N. T. Drane found, for example, this U-shaped phenomenon for both the entire federal and the state-local structures in Australia. In two separate studies, W. Irwin Gillespie and Allan Maslove discovered the same pattern in Canada, as did Pechman and Okner for taxes in the United States. Clearly the S index will be an unreliable indicator in these broad, varying, and pervasive cases, and it is unlikely, even if carefully applied, that it will be a "...widely useful measure of tax progressivity..." (Suits, p. 752).

Moreover, the S index does not reveal information about the skewness in the dis-

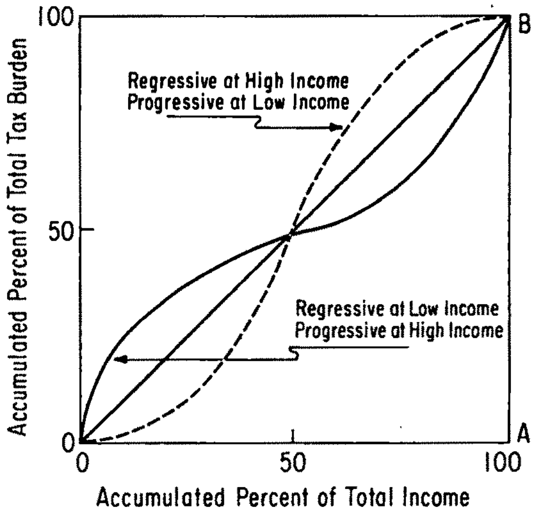


FIGURE 2

[6]For rigorous and excellent discussions of the enormously complicated problems associated with attempts to measure and compare with precision alternative distributions of income, see Amartya Sen and Martin Bronfenbrenner, chs. 1–5.

tribution of income nor the number of people or families subject to progressivity, and this lack of evidence could cause the policymaker difficulty. For example, we know from the statistical data Suits presents in his Table 1 that the first eight population deciles in the United States received approximately 50 percent of the income in 1966. We also know that the S index of tax B (curved solid line) in Suits' Figure 5 (my Figure 2) has a value of 0 which causes Suits to label it a proportional tax. Yet if the distribution of income underlying Figure 5 were like that of the United States for 1966, approximately 80 percent of the families would fall in the regressive zone. Conversely Tax A (dotted line), which Suits also labels proportional, would have the lowest 80 percent of the families subject to a tax progression and the highest 20 percent to tax regression. A more useful index would have to incorporate this information on number of families some way; a policymaker whose decisions were based on the S index would otherwise be confronted with a politically explosive situation. Until such a useful index is formulated, circumstances call for disaggregation of the data. Nevertheless, despite these criticisms, the profession is in Suits' debt for pushing forward with his ingenious index.

REFERENCES

A. Ando and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *Amer. Econ. Rev.*, Mar. 1963, *53*, 55–84.

Henry J. Aaron, *Who Pays the Property Tax?*, Washington 1975.

P. Bently, D. Collins, and N. Drane, "An Estimate of the Incidence of the Australian Tax Structure," paper presented at Economic Society of Australia and New Zealand, Adelaide 1973.

Martin Bronfenbrenner, *Income Distribution Theory*, Chicago 1971.

Edgar K. Browning, "The Trend Toward Equality in the Distribution of Net Income," *Southern Econ. J.*, July 1976, *43*, 912–923.

Joseph A. Califano, Jr., "Remarks," *HEW News*, April 27, 1977, 1–9.

David G. Davies, "A Further Reappraisal of Sales Taxation," *Nat. Tax J.*, Dec. 1963, *16*, 410–15.

______, "An Empirical Test of Sales-Tax Regressivity," *J. Polit. Econ.*, Feb. 1959, *67*, 72–78.

______, "Clothing Exemptions and Sales Tax Regressivity: Note," *Amer. Econ. Rev.*, Mar. 1971, *61*, 187–189.

______, "Commodity Taxation and Equity," *J. Finance*, Dec. 1961, *16*, 581–90.

______, *International Comparisons of Tax Structures in Federal and Unitary Countries*, Canberra 1976.

______, "Progressiveness of a Sales Tax in Relation to Various Income Bases," *Amer. Econ. Rev.*, Dec. 1960, *50*, 987–95.

W. Irwin Gillespie, *The Incidence of Taxes and Public Expenditures in the Canadian Economy*, Ottawa 1967.

John J. Korbel, *Poverty Status of Families Under Alternative Definitions of Income*, Congressional Budget Office, U.S. Congress, Washington 1977.

Allan M. Maslove, *The Pattern of Taxation in Canada*, Ottawa 1972.

T. Mayer, "The Distribution of the Tax Burden and Permanent Income," *Nat. Tax J.*, March 1974, *27*, 141–146.

Richard A. Musgrave, *The Theory of Public Finance*, New York 1959, 224–25.

______ **and T. Thin,** "Income Tax Progression, 1929–48," *J. Polit. Econ.*, Dec. 1948, *56*, 498–514.

Joseph A. Pechman and Benjamin A. Okner, *Who Bears the Tax Burden?*, Washington 1974.

Morgan Reynolds and Eugene Smolensky, *Public Expenditures, Taxes, and the Distribution of Income*, New York 1977.

J. M. Schaefer, "Clothing Exemptions and Sales Tax Regressivity," *Amer. Econ. Rev.*, Sept. 1969, *59*, 596–99.

Amartya Sen, *On Economic Inequality*, Oxford 1973.

D. B. Suits, "Measurement of Tax Progressivity," *Amer. Econ. Rev.*, Sept. 1977, *67*, 747–52.

L. C. Thurow, "The Economics of Public Finance," *Nat. Tax J.*, June 1975, *28*, 185–94.

Measurement of Tax Progressivity: Comment

By EDWARD C. KIENZLE*

In a recent issue of this *Review*, Daniel Suits contributed an excellent method for deriving a measure of tax progressivity (or regressivity). The "Suits Index" S, closely related to Lorenz curve and Gini coefficient analysis, is an ingenious extension of that measurement methodology. It provides a widely applicable summary statistic of the progressivity of a particular tax or tax structure.[1]

Suits demonstrates the mathematical properties of S and applies the index to the *U.S.* tax structure. He examines 1966 tax data developed by Joseph Pechman and Benjamin Okner, and 1970 tax data developed by Okner (1976) and compares the results for the two years.

Economic theory does not provide a definitive answer to tax incidence questions and the subsequent allocation of tax burdens across income classes. Application of legitimate competing theories concerning tax incidence results in different implications for tax progressivity. Therefore, alternative assumptions should be considered when measuring tax progressivity. The purpose of this paper is to examine the results obtained from an application of the Suits Index and Lorenz curves of taxes when an alternative set of assumptions is chosen and to show that S and the Lorenz curves are sensitive to that choice.

I. Results

In analyzing the distribution of tax burdens by income classes, Pechman and Okner make varying assumptions concerning tax incidence.[2] Suits' empirical results for 1966 are based entirely on variant 1c, described by Pechman and Okner as the most progressive set of incidence assumptions utilized. For purposes of comparison, this investigation uses variant 3b, classified by Pechman and Okner as the least progressive set.[3]

The data appearing in Table 1 are analogous to Suits' Table 1. The first column gives accumulated percentages of families in order of income, grouped by deciles, and the second column gives the associated accumulated percentages of total adjusted family income.[4] The remaining columns give the accumulated percentages of total tax revenue contributed by the corresponding accumulated deciles of families. For example, the fourth line of the table, associated with the lowest 40 percent of families in the income distribution, reveals that these accumulated families received 14.15 percent of total adjusted family income, contributed 6.43 percent of total individual income tax revenue, contributed 15.79 percent of the corporate income tax, and so on. Table 2 contains two columns of the calculated progressivity indexes for 1966 under 1c and 3b as discussed above.

Since the Suits Index itself may fail to reveal significant differences in Lorenz curves of taxes under different tax incidence assumptions, an investigation of movements both in Lorenz curves of taxes and in the Suits Index is necessary for a complete analysis of the sensitivity of tax progressivity to alternative tax incidence assumptions. Figure 1 illustrates Lorenz curves for the corporate income tax and property taxes using 3b. These Lorenz curves are changed substantially from what thay are under 1c, but the

*Assistant professor, department of economics, Central Connecticut State College. I am grateful to an anonymous referee for helpful comments.

[1]In fact, Suits has already published additional applications of S (1977a).

[2]These sets of assumptions are structured into three broad "variants" which are then subdivided, giving a total of eight variants. The major differences among the variants are related to the presumed incidence of the property, corporation, and payroll taxes (see Pechman and Okner, p. 37).

[3]For a discussion of the specific incidence assumptions with respect to the alternative variants, see Pechman and Okner (pp. 25–43).

[4]See Pechman and Okner (Tables 4–6, p. 56).

TABLE 1—ACCUMULATED *U.S.* INCOME AND TAX BURDEN BY POPULATION DECILES, 1966

Cumulated Percentage: Population Decile	Adjusted Family Income	Individual Income Tax	Corporate Income Tax	Property Taxes	Sales and Excise Taxes	Payroll Taxes	Personal Property and Motor Vehicle Taxes	All Federal Taxes	All State and Local Taxes	All Taxes
1	1.25	0.18	1.73	2.42	2.22	1.30	1.80	0.97	2.18	1.34
2	3.98	0.83	5.07	6.64	6.31	4.13	5.74	3.07	6.05	3.98
3	8.29	2.85	9.95	12.65	12.16	9.49	11.96	6.90	11.66	8.35
4	14.15	6.43	15.79	19.39	19.76	17.19	18.30	12.43	18.46	14.27
5	21.56	11.76	22.67	26.80	28.93	27.09	26.32	19.68	26.40	21.72
6	30.85	19.22	30.87	35.80	39.87	39.08	36.37	28.77	36.12	30.99
7	41.01	28.11	39.38	45.65	51.23	51.72	47.37	38.77	46.49	41.09
8	53.44	40.04	49.79	57.70	64.38	66.04	60.82	51.14	59.02	53.45
9	68.92	55.63	63.45	72.70	79.52	82.09	77.57	66.46	73.83	68.60
10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Addendum: Average Tax Rate		8.4	4.4	3.4	5.0	4.4	0.3	17.6	7.6	25.9

Source: Calculated from data in Pechman and Okner using variant 3b.

TABLE 2—PROGRESSIVITY OF *U.S.* TAXES UNDER ALTERNATIVE TAX INCIDENCE ASSUMPTIONS, 1966

	Suits Index S Variant 1c	Suits Index S Variant 3b
Individual income tax	.17	.186
Corporate income tax	.36	.033
Property taxes	.23	−.072
Sales and excise taxes	−.16	−.146
Payroll taxes	−.17	−.151
Personal property and motor vehicle taxes	−.12	−.105
All federal taxes	.087	.035
All state and local taxes	.045	−.080
All taxes	.074	.0005

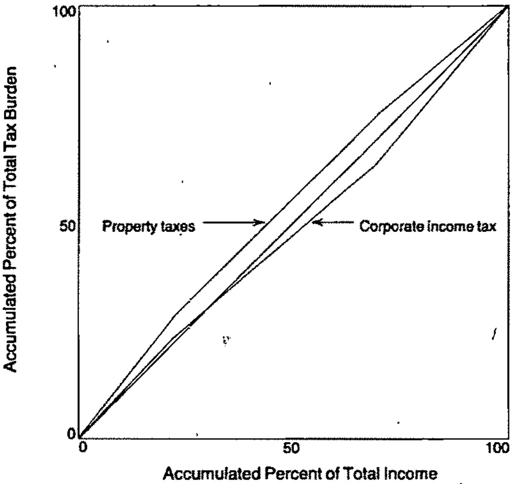


FIGURE 1. LORENZ CURVES FOR TWO *U.S.* TAXES

Lorenz curves for the four other disaggregated taxes (individual income, sales and excise, payroll, and personal property and motor vehicle) are not changed substantially under 3b compared to 1c.

II. Conclusions

The results indicate that S is indeed sensitive to the choice of an alternative set of assumptions, in this case, 3b. The greatest change toward regressivity occurs in the property taxes and in the corporate income tax.[5] Examination of S indicates that the

[5]See Suits (1977b, p. 749) for a discussion of the range of S. The Suits Index is positive for progressive taxes and in the limiting extreme case where the entire

property tax under 3b is regressive; whereas it is progressive under 1c.[6] Moreover, with respect to the Lorenz curves of taxes, the property taxes are progressive throughout the entire range of incomes under 1c, but Table 1 and Figure 1 reveal them to be regressive throughout under 3b. The index S measures the corporate income tax as substantially progressive under 1c, but only slightly progressive under 3b.[7] While this tax is continuously progressive under 1c, Table 1 and Figure 1 reveal that under 3b it is regressive over the lower six deciles of families and progressive beyond.

As a whole, the federal tax structure remains progressive ($S=.035$) even under the least progressive assumptions, but the state and local tax structure becomes regressive ($S=-.08$). Overall, the total tax structure remains progressive but only slightly so ($S=.0005$).[8] Thus, the substantially more regressive set of assumptions (3b) concerning property taxes makes the total state and local tax structure regressive, and this, in conjunction with the more regressive assumptions concerning the corporate income tax, tips the total tax structure toward regressivity. Although this shift does not result in overall regressivity, measured progressivity is so slight that the total tax structure might be termed proportional under 3b.

tax burden is borne by the highest-income group, $S=+1$. For a proportional tax, $S=0$, and for a regressive tax S is negative. In the limiting extreme case for a regressive tax, the entire tax burden is borne by the lowest income group and $S=-1$. Hence, $-1 \leq S \leq +1$, and the more positive S becomes (or the less negative), the more progressive (or less regressive) the tax becomes.

[6]Under 1c, Pechman and Okner regard the property tax on improvements and on land as taxes on property income in general to be distributed in proportion to the property income reported by each household (pp. 37–38); while under 3b, they assume that the property tax on land is capitalized and borne by landowners in proportion to the value of owned land and that recipients of property income bear only half of the property tax on improvements, the other half being shifted in the form of higher shelter and consumption good prices (pp. 37–39).

[7]Under 1c, half of the corporation income tax is borne by property owners in general and the other half by stockholders (Pechman and Okner, p. 38); while under 3b, half of the corporate income tax is again assumed to be borne by property owners in general, but the other half is borne by consumers (p. 39).

[8]This result conflicts with the assertion of Pechman and Okner that 3b "...produces a slightly regressive distribution" (p. 57). It can be seen in Table 1, however, that the category "All Taxes" is indeed slightly regressive over all but one of the lower eight deciles of families.

REFERENCES

B. A. Okner "Total U.S. Tax Burdens: 1966 and 1970 Compared," paper presented at the DFG Symposium, Augsburg, Germany, July 1976.

Joseph A. Pechman and Benjamin A. Okner, *Who Bears the Tax Burden?*, Washington 1974.

D. B. Suits, (1977a) "Gambling Taxes: Regressivity and Revenue Potential," *Nat. Tax J.*, Mar. 1977, *30*, 19–35.

______, (1977b) "Measurement of Tax Progressivity," *Amer. Econ. Rev.*, Sept. 1977, *67*, 747–52.

Measurement of Tax Progressivity: Reply

By DANIEL B. SUITS*

Two quite different issues are raised by the comments of David Davies and Edward Kienzle. The first, raised by both authors, is that use of *S* (or, for that matter, any other index of tax progressivity) in no way avoids fundamental problems of tax shifting, questions of how income should be measured for proper assessment of tax burden, or how to treat the difference between the distribution of lifetime tax burden compared to what is observed in any given year. The second issue, raised by Davies, deals with the validity of the index *S*, itself.

Estimates of tax incidence differ with the kind of assumptions made about tax shifting. In this regard, the greatest variation occurs among estimates of the distribution of the burden of taxes on property, including the corporate income tax. Kienzle demonstrates this in an interesting way in his Table 2. Whether variant 1c or 3b is employed makes little difference in the value of *S* calculated for income, sales, or payroll taxes, but taxes on property appear highly progressive under 1c, and virtually proportional under 3b.

Davies makes a similar point in noting that the value of *S* will depend on how income is measured. It would be interesting to see what difference would be produced if the adjustments to income he suggests were made. Likewise, it would be useful to see what happens to *S* when both income and tax burden are estimated on a lifetime basis.

*The East-West Population Institute, Hawaii, and Michigan State University.

I have made a few attempts to make such calculations, but have so far been defeated by the problems of keeping track of who the taxpayer is, especially with the occurrence of divorce, widowhood, and remarriage.

Davies also raises questions about the validity of *S* itself. His first objection—that no single index can summarize all aspects of a complicated phenomenon—largely duplicates what I said in the original presentation. The limitation is common to all averages and indexes (think of the divergence of individual price behavior concealed in movements of the Consumer Price Index) but we still find them useful.

In his concluding remarks, Davies comes close to saying that tax incidence is too complex to permit tax progressivity to be measured at all. If this were true, of course, neither *S* nor any other such index would make sense. Personally, however, I think some useful things can be said about the relationship of tax burden to income, and that *S* serves as a good summary measure of those things.

REFERENCES

D. G. Davies, "Measurement of Tax Progressivity: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 204–07.

E. C. Kienzle, "Measurement of Tax Progressivity: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 208–10.

D. B. Suits, "Measurement of Tax Progressivity," *Amer. Econ. Rev.*, Sept. 1977, *67*, 747–52.

Price Uncertainty and the Cooperative Firm

By JACOB PAROUSH AND NAVA KAHANA*

The behavior of a cooperative firm which is interested in the well-being of its members has attracted economists' attention in quite a few studies. The pioneer work of Benjamin Ward compares the policy of such a firm with that of a profit maximizer under perfect competition. The analysis is further carried out by Evsey Domar (1966) and Jaroslav Vanek (1969). The comprehensive study of Vanek (1970) establishes a general theory of labor-managed market economies. The theory of cooperatives and participatory structures of enterprises has recently been further extended in several directions by S. Charles Maurice and Charles E. Ferguson, J. E. Meade, Eirik Furubotn, among others. However, the basic assumption in all these studies is that the cooperative operates with complete information about the prices of the final demand markets.

This paper is concerned with the behavior of a competitive cooperative firm under price uncertainty. That is, the price of the final product is not perfectly anticipated, for instance, as a result of inflation, but is rather uncertain and takes the form of a stochastic variable with a known distribution. A recent paper of Allan Taub deals with the very same subject. However, the treatment here is more comprehensive and the results are entirely different.[1]

Compare the output policy of two types of firms: The profit maximizer whose labor input consists of workers with no share in the firm's profits and a second firm which seeks the optimum number of members to maximize profits or net income per member. We shall call the first type a capitalist and the second type a cooperative. As noted by Vanek (1970) and others, there are several aspects which distinguish the cooperative firm from the capitalist firm. However, we disregard all the other characteristics and concentrate in this paper on the difference of the objective functions of the two types of firms.

Ward has already shown that under perfect competition and with complete price anticipation, a cooperative produces less than a capitalist who has the same production technology. On the other hand, Agnar Sandmo and Hayne Leland have found that under price uncertainty a risk-averse capitalist firm produces less than under complete certainty. The main finding of this paper is that Sandmo's result does not apply to cooperatives. That is, a cooperative even if risk averse produces *more* under uncertainty and therefore increases its demand for labor input. As a consequence of this result, Ward's conclusion is weakened and becomes less significant. That is, in the case of price uncertainty the discrepancy between cooperative and capitalist with respect to production and occupation is smaller than under price stability. Section I states the assumptions of the model and compares the optimum solution of a capitalist with that of a cooperative firm under certainty as well as under risk conditions. Section II presents comparative statics first with respect to the basic parameters of the price distribution (the average price and the variance), and second, with respect to the basic parameters of the profit function (the wage rate and the fixed cost).

I. A Comparison of Optimum Solutions

Assume a firm which produces a homogeneous product y with a single variable production factor x. The variable factor x is the labor input, and is taken in the following analysis as a continuous variable. The production function is assumed to display di-

*Bar-Ilan University, Ramat-Gan, Israel. We acknowledge the helpful comments made by two anonymous referees.

[1]Taub presents his arguments in a way that is impossible to follow. Coming up with different results we suspect some unstated assumptions.

minishing returns to labor input, i.e.,

$$(1)\qquad y=f(x),\quad f'>0,\quad f''<0$$

for all x in the relevant range. Note that the one-to-one correspondence between input and output allows us to draw inference about optimum outputs from comparison of optimum inputs under the several circumstances analyzed. We entirely disregard the important and nontrivial problem of comparing the optimum capital levels in the long run under the different conditions analyzed here by avoiding the use of a many-variables production function.[2]

The objective functions of a capitalist and that of a cooperative are defined, respectively, as

$$\pi_0=pf-wx-F \quad \text{and}$$

$$\pi_c=(pf-wx-F)/x=(pf-F)/x-w$$

Here, p is the unit price of output, w is the wage rate per labor unit, and F is a fixed cost paid for the fixed factors. Under certainty the price of the final product is assumed to be fixed and known, say $p=\bar{p}$, and the target function is a utility function defined over profit or profit per member, $u(\pi)$ that has to be maximized with respect to x. We assume that the firm or the decision maker is a risk averter so that u is a concave function,[3] i.e.,

$$(2)\qquad u'>0,\quad u''<0$$

[2]Note that in this paper all our results concerning output rely heavily on the one-to-one correspondence between employment and production assumed in (1). In an unpublished dissertation written by Kahana under the supervision of Paroush, the present analysis to the case of many production factors is extended. In the general case we specify exactly the conditions under which some of the present results still hold. These conditions turn out to be quite weak and, as one may expect, have to do with the substitution (the sign of the cross derivative) between labor and capital.

[3]In the case of the capitalist firm the utility function is possessed by the owner. In the case of the labor-managed firm the utility is a "group" utility function. In order to avoid aggregation problems one may assume that each member of the cooperative has the same utility function with the same attitude toward risk.

Under uncertainty p is a stochastic variable with $\bar{p}$ and γ as the expected value and the standard deviation, respectively. The target function in this case is the expected utility suggested by John von Neumann and Oskar Morgenstern which has to be maximized with respect to the decision variable x.

We assume that the four optimization problems $\operatorname{Max} u(\pi_0)$, $\operatorname{Max} u(\pi_c)$, $\operatorname{Max} Eu(\pi_0)$, and $\operatorname{Max} Eu(\pi_c)$ have regular, unique, and interior solutions in the positive relevant range of x. Denote these optimum points as x_0, x_c, x_0^*, and x_c^*, respectively. That is, the first- and the second-order conditions are assumed to hold at these points.

Sandmo (p. 66) finds that a capitalistic firm produces less under price uncertainty than its certainty equivalent twin, i.e., $x_0>x_0^*$. Here we prove that Sandmo's result does not hold for a cooperative. The cooperative firm, even if risk averse, produces *more* and increases its demand for labor input under price fluctuations than its certainty equivalent twin, i.e.,

PROPOSITION 1: *Under equations* (1) *and* (2), $x_c<x_c^*$.

PROOF:

Substitute $d\pi_c/dx=[p(xf'-f)+F]/x^2$ into the first-order condition $Eu'(d\pi_c/dx)=0$ to find that at the point x_c^*

$$Cov[u',p(xf'-f)+F]=-E[p(xf'-f)+F]Eu'$$

or

$$(3)\qquad \bar{p}(xf'-f)+F=-(xf'-f)Cov(u',p)/Eu'$$

where $\bar{p}=Ep$. In (3), $Eu'>0$ and $Cov(u',p)$ has the sign of $du'/dp=u''(d\pi_c/dp)$ which is negative under (2). Therefore, at x_c^* the left-hand side of (3) has the sign of $xf'-f$ which is negative at the optimum points. Under certainty $\bar{p}(xf'-f)+F=0$ at x_c and decreases with x. Consequently $x_c^*>x_c$.

In his pioneer study Ward (p. 576) observes that a cooperative employs a smaller number of workers than its capitalist twin. The main important corollary of Proposition 1 is that Ward's effect loses much of its significance under price uncertainty. First, we shall prove this corollary and then we shall briefly discuss its applications.

Denote by $\Delta x = x_0 - x_c$ and $\Delta x^* = x_0^* - x_c^*$ the Ward effects under fixed price and under price uncertainty, respectively.

COROLLARY 1: $\Delta x^* < \Delta x$

PROOF:

Sum up Sandmo's result: $x_0 - x_0^* > 0$ together with that of Proposition 1, $x_c^* - x_c > 0$, to find that

$$(x_0 - x_0^*) + (x_c^* - x_c) = (x_0 - x_c) - (x_0^* - x_c^*) > 0$$

Two comments are in order. First, note that although Ward finds that $\Delta x > 0$ one cannot state in general that $\Delta x^* > 0$. It is quite possible that under price uncertainty, a cooperative may produce even more than its capitalist twin. However, if risk is small enough, then $\Delta x^* > 0$ holds. Risk is small enough if the distances $x_c^* - x_c$ and $x_0 - x_0^*$ are of a smaller order of magnitude than the finite and fixed difference $x_0 - x_c$. Second, relying on Ward's result, governments, boards of planning, or regulatory authorities may object and restrict any cooperative movement and any economic activity of labor-managed firms. The alleged argument may be that such organizations decrease output as well as occupation and therefore reduce total welfare, especially in periods of recessions and unemployment. As a result of our analysis such a policy is not correct. Under price uncertainty in general, and under inflation in particular, cooperatives may produce more than capitalists and may have a higher labor derived demand. Even if inflation or price fluctuations are modest such that Ward's effect is still positive, it may be quite negligible in comparison to the effect under certainty.

II. A Comparative Statics Analysis

The comparative statics analysis performed here is with respect to the parameters of the price distribution. Following Sandmo, two parameters are considered, the average price $\bar{p}$ and the variance of p, denoted by γ. However, one has to keep in mind that the variance of p is a measure of risk under one of the following three conditions: 1) either the utility function is quadratic, 2) the price distribution is normal, or 3) risk is "small." (See Paul Samuelson.) Recall that the firms considered here operate under perfect competition in the sense that they are price takers. That is the output prices in the case of complete foresight or the price distribution in the case of uncertainty are given to the decision maker as external information. So any price change may be taken as a result of an exogeneous change of total demand.

The first result to be proved is that the reduction of output produced by the cooperative in comparison to its capitalist twin is larger the higher is the demand for the final product. Moreover, the above result holds under price uncertainty as well as under perfect price anticipation. More specifically, we shall prove that $d\Delta x / d\bar{p} > 0$ and $d\Delta x_c^* / d\bar{p} > 0$. However, another assumption has to be added to the model and that is a nonincreasing absolute risk aversion[4] displayed by the decision maker, i.e.,

$$(4) \quad -u''(\pi_2)/u'(\pi_2) \leqslant -u''(\pi_1)/u'(\pi_1) \quad \text{if and only if } \pi_2 > \pi_1$$

PROPOSITION 2: *Under equations* (1), (2), *and* (4), $dx_c^*/d\bar{p} < 0$ *and* $dx_c^*/d\gamma > 0$.

PROOF:

An implicit differentiation of the first-order condition yields

$$dx_c^*/d\bar{p} = -\left(d^2Eu(\pi_c)/dx d\bar{p}\right)/\left(d^2Eu(\pi_c)/dx^2\right)$$

[4] On the measurement and applications of risk aversion see Kenneth Arrow and John Pratt.

Use the second-order condition to find that

$$Sgn(dx_c^*/d\bar{p}) = Sgn(d^2Eu(\pi_c)/dx\,d\bar{p}) \text{ at } x_c^*$$

By differentiation one can find that

$$Sgn\left[d^2Eu(\pi_c)/dx\,d\bar{p}\right]$$

$$= Sgn\{Eu''\left[p(xf'-f)+F\right]+Eu'(xf'-f)\}$$

$$= Sgn(xf'-f)\{Eu''\left[p+F/(xf'-f)\right]+Eu'\}$$

Sandmo (pp. 68–89) shows that $Eu''[p+F/(xf'-f)] \gtreqless 0$ under decreasing, constant, or increasing absolute risk aversion, respectively. Since $Eu'>0$, and $(xf'-f)<0$ the sign of the above expression is negative which proves that $dx_c^*/d\bar{p}<0$.

To prove that $dx_c^*/d\gamma>0$, one can follow Yasunori Ishii's paper which is based on Sandmo's technique of "mean preserving" change in risk, noting that the sign reversal is caused by the negative term $(xf'-x)$ as above.

The following corollary summarizes the effects of $\bar{p}$ and γ on Ward's effects.

COROLLARY 2: *Under equations* (1), (2), *and* (4), $d\Delta x^*/d\bar{p}>0$ *as well as* $d\Delta x/d\bar{p}>0$ *but* $d\Delta x^*/d\gamma<0$.

PROOF:

$dx_0/d\bar{p}>0$	by standard microeconomics
$dx_c/d\bar{p}<0$	is shown by Ward (p. 576)
$dx_0^*/d\bar{p}>0$	is derived by Sandmo (p. 69)

Combine these three results with $dx_c^*/d\bar{p}<0$ in Proposition 2 to find that $d\Delta x^*/d\bar{p}>0$ as well as $d\Delta x/d\bar{p}>0$. Ishii (pp. 768–69) shows that $dx_0^*/d\gamma<0$. Combine this result with $dx_c^*/d\gamma>0$ in Proposition 2 to find that $d\Delta x^*/d\gamma<0$.

One can interpret the changes $d\bar{p}$ and $d\gamma$ as an anticipated and unanticipated inflation, respectively. The results $d\Delta x^*/d\bar{p}>0$ and $d\Delta x^*/d\bar{p}>0$ mean that Ward's effect increases with anticipated inflation, and the result $d\Delta x^*/d\gamma<0$ means that Ward's effect loses its significance with unanticipated inflation.[5] This is compatible with our finding in Section I that Ward's effect is smaller under price uncertainty than under price certainty, i.e., $\Delta x^*<\Delta x$.

We shall complete this section with a static comparison with respect to F and w. From standard microeconomics one knows that $dx_0/dF=0$, $dx_0/dw<0$. Ward (p. 566) shows that $dx_c/dF>0$ and $dx_c/dw=0$. One of the interesting results of Sandmo (p. 65) is that $dx_0^*/dF \leqslant 0$ and a direct result from the analysis of Raveendra Batra and Aman Ullah (p. 537) is that $dx_0^*/dw<0$. We next prove that in contrast to these cases, $dx_c^*/dF>0$ and $dx_c^*/dw \geqslant 0$. In other words, assuming nonincreasing absolute risk aversion, the cooperative always increases optimum output and labor input as a result of an increase of the fixed cost regardless of the risk conditions, while the capitalist would never do it.

PROPOSITION 3. *Under equations* (1), (2), *and* (4), $dx_c^*/dF>0$ *and* $dx_c^*/dw \geqslant 0$.

PROOF:

As before, implicitly differentiate the first-order condition and use the second-order condition to find that at the optimum

$$Sgn(dx_c^*/dF) = Sgn(d^2Eu(\pi_c)/dx\,dF)$$

$$Sgn(dx_c^*/dw) = Sgn(d^2Eu(\pi_c)/dx\,dw)$$

Differentiate the right-hand side of the above equations to find that

$$Sgn(dx_c^*/dF) = Sgn(-Eu''v+Eu')$$

$$Sgn(dx_c^*/dw) = Sgn(-Eu''v)$$

where $v=p(xf'-f)+F$. Following the analysis of Sandmo one can find that under (4) $Eu''v \leqslant 0$ which completes the proof.

[5]The association of $d\bar{p}$ and $d\gamma$ with anticipated and unanticipated inflation is more realistic in a situation where there is some time lag between present decision about labor input with a constant w and future realization of the price of the final product.

It is worthwhile to conclude this paper with a brief intuitive explanation of the somewhat surprising results which emerge from our analysis. The basic question is why the cooperative responds to price uncertainty in opposite fashion to the capitalist, even though both firms have the same attitude toward risk. It seems that the key point for understanding the answer to this question is the relation between price and optimum output. The capitalist's supply function has a positive slope, while the cooperative, as shown by Ward, has a downward-sloping supply function. More specifically, the capitalist firm responds to a price rise in the natural way by increasing optimum output so that $dx_0/d\bar{p}>0$ and $dx_0^*/d\bar{p}>0$. The cooperative, on the other hand, has a negative response so that $dx_c/d\bar{p}<0$ and $dx_c^*/d\bar{p}<0$. Since the utility function is assumed to be concave and since profits as well as profit per labor unit are linear functions of price, the indirect utility function is also concave in the price $\bar{p}$. The introduction of uncertainty into the system has therefore exactly the same impact as a price decline. In other words, the certainty equivalent price, say $\hat{p}$, is smaller than $\bar{p}$ because of Jensen's inequality, i.e., $u(\hat{p})=Eu(p)<u(\bar{p})$. This may serve as an intuitive explanation of why the two kinds of firms behave differently under price uncertainty and $x_0^*<x_0$ while $x_c^*>x_c$.

REFERENCES

Kenneth J. Arrow, *Essays in the Theory of Risk Bearing*, Amsterdam 1970, ch. 3.

R. N. Batra and A. Ullah, "Competitive Firm and the Theory of Input Demand Under Price Uncertainty," *J. Polit. Econ.*, May/June 1974, *82*, 537–48.

E. Domar, "The Soviet Collective Farm as Producer Cooperative," *Amer. Econ. Rev.*, Sept. 1966, *56*, 734–57.

E. G. Furubotn, "The Long-Run Analysis of the Labor-Managed Firm: An Alternative Interpretation," *Amer. Econ. Rev.*, Mar. 1976, *66*, 104–23.

Y. Ishii, "On the Theory of the Competitive Firm under Price Uncertainty: Note," *Amer. Econ. Rev.*, Sept. 1977, *67*, 768–69.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

S. C. Maurice and C. E. Ferguson, "Factor Usage by a Labor-Managed Firm in a Socialist Economy," *Economica*, Feb. 1972, *39*, 18–31.

J. E. Meade, "Labour-Managed Firms in Conditions of Imperfect Competition," *Econ. J.*, Dec. 1974, *84*, 817–24.

J. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan./Apr. 1964, *32*, 122–36.

P. A. Samuelson, "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variance and Higher Moments," *Rev. Econ. Stud.*, Oct. 1970, *37*, 537–42.

A. Sandmo, "On the Theory of the Competitive Firm under Price Uncertainty," *Amer. Econ. Rev.*, Mar. 1971, *61*, 65–73.

A. J. Taub, "The Cooperative Firm under Price Uncertainty," *Eastern Econ. J.*, Apr. 1974, *1*, 167–69.

Jaroslav Vanek, "Decentralization Under Workers' Management: A Theoretical Appraisal," *Amer. Econ. Rev.*, Dec. 1969, *59*, 1006–14.

______, *The General Theory of Labor-Managed Market Economies*, Ithaca 1970.

John von Neumann and Oscar Morgenstern, *Theory of Games and Economic Behavior*, 3d ed., Princeton 1953, ch. I, sec. 3.

B. Ward, "The Firm in Illyria: Market Syndicalism," *Amer. Econ. Rev.*, Sept. 1958, *48*, 566–89.

The Ranking of Behavioral Modes of the Firm Facing Uncertain Demand

By CHIN LIM*

The theory of the firm under uncertainty has been subject to a variety of studies but most of the analyses have restricted the firm to a singular mode of behavior. Little attention has been given to the fact that uncertainty enables the firm to choose from among various behavioral modes. The recognition of this problem was first due to David Baron and Hayne Leland who observed that under uncertain demand conditions, feasible behavioral modes could be classified according to whether price and/or quantity decisions are made prior to the revelation of the true demand conditions. Thus, before the true demand conditions are revealed, the firm could either decide on 1) quantity (henceforth called Q behavior), or 2) price (henceforth called P behavior), or 3) quantity-cum-price (henceforth called Q/P behavior).[1] The natural question that follows—which I address in this paper—is how the firm would rank these behavioral modes if it is able to choose among them. The results from this analysis are useful in providing valuable insight as to why firms choose to behave in a particular way in the market.

Previous attempts at comparison of the above behavioral modes have been unsatisfactory. For instance, Leland noted generally that no a priori judgement can be made about the relative dominance between Q and P behavior, but both these behavioral modes should dominate Q/P behavior. This is because Q and P behavior are argued to be more "flexible"[2] than Q/P behavior. Such a conclusion is generally correct provided that the profit functional forms are identical in all the behavioral modes. Unfortunately this is not so and the ranking of behavioral modes cannot simply depend on the *ex post* flexibility criterion. Another attempt, this time due to Baron, did not successfully result in a complete ranking of the various behavioral modes. This is because the comparison of modes was not evaluated for different cost conditions; moreover, different specific examples regarding demand conditions were used for different pairwise comparisons of behavioral modes.

In this paper, I provide a common framework for the analysis of relative dominance among the various behavioral modes. Section I contains a characterization of the modes. Section II analyzes their relative dominance. Finally, a summary of results and their implications is presented in Section III.

*Assistant professor, University of Western Ontario. This paper is based on chapter 2 of my doctoral dissertation submitted to Queen's University. I am grateful to my thesis committee, R. G. Harris and F. R. Flatters, and also to D. Purvis, H. Oniki, P. Howitt, A. Ullah, and an anonymous referee for helpful comments. Without implicating them in any particular view, all errors are mine alone.

[1]The ability to set price presumes the firm faces a downward-sloping demand function. However, this need not imply the firm is a monopolist in the classical sense of the word; that is, there exists only a single firm in the market. It is well known that if consumers face imperfect information and have to incur costly search to discover the distribution of prices charged elsewhere, then even a competitive firm can behave like a monopolist. A similar conclusion is observed by Kenneth Arrow, but stems from a different reason. Arrow noted that the assumption of a competitive firm being a price taker is only reasonable if the price is in fact an equilibrium price. Otherwise, at disequilibrium prices, a competitive firm can temporarily behave as a monopolist.

[2]The general argument is as follows: In Q behavior, the firm produces prior to the revelation of the true demand conditions, but is able to set price *ex post*. In P behavior, the price is set prior to the revelation of the true demand conditions, but production decision is made *ex post*. Thus, both P and Q behavior are considered more *ex post* flexible when compared to Q/P behavior where the price and quantity decision is made *ex ante*, leaving no instrument of control *ex post*.

I. Characterization of Behavioral Modes

The firm is assumed to face a random demand function. Following Leland, the stochastic demand relation is expressed in most general form as the implicit demand function,

$$(1) \qquad F(q,p,\eta)=0$$

where q is quantity, p is price, and $\eta \in [\underline{M}, \overline{M}]$ is a random variable which is not known *ex ante* but has subjective probability density $d\Psi(\eta)$. Treating (1) as a mapping from a probability space into a set of downward-sloping demand functions, we can place the following restrictions on the implicit form. For any given η, let the relation between p and q be downward sloping, so that larger values of η are associated with greater demand. These restrictions allow us to express (1) as either[3]

$$(2) \quad q=q(p,\eta), \quad q_1(p,\eta)<0, \quad q_2(p,\eta)>0$$

or

$$(3) \quad p=p(q,\eta), \quad p_1(q,\eta)<0, \quad p_2(q,\eta)>0$$

Now, define $\eta^o \equiv E[\eta]$ where E is an expectations operator, and consider two arbitrary values of p_c and q_c. For sufficiently concentrated distributions, we can accurately approximate $q(p_c,\eta)$ and $p(q_c,\eta)$ around η^o as

(4)

$$q(p_c,\eta)\simeq q(p_c,\eta^o)+q_2(p_c,\eta^o)\cdot[\eta-\eta^o]$$

$$p(q_c,\eta)\simeq p(q_c,\eta^o)+p_2(q_c,\eta^o)\cdot[\eta-\eta^o]$$

Taking the expectation of the above, and noting that $\eta^o = E[\eta]$, we obtain

$$(5) \qquad E[q(p_c,\eta)]\simeq q(p_c,\eta^o)$$

$$E[p(q_c,\eta)]\simeq p(q_c,\eta^o)$$

The above linear approximation of the random demand function is accurate provided (see Paul Samuelson) the probability distribution is compact and concentrated (i.e., for small uncertainty or random fluctuation in η). My ensuing theorems must therefore be carefully interpreted to apply to sufficiently small uncertainty.[4]

Now, because $q_1(p,\eta)$, $p_1(q,\eta)<0$, it immediately follows that for any arbitrary choice of q_c and p_c,[5]

$$q_c \lesseqgtr q(p_c,\eta^o) \Leftrightarrow p_c \lesseqgtr p(q_c,\eta^o)$$

which, after using (5), implies that for small variations in η,

$$(6) \quad q_c \lesseqgtr E[q(p_c,\eta)] \Leftrightarrow p_c \lesseqgtr E[p(q_c,\eta)]$$

On the cost side, the firm is assumed to face, with certainty,[6] a cost function

$$(7) \qquad c=c(q), \qquad c'(q)>0$$

where q is output, and $c'(q)$ is the marginal cost.

With the above information set, the firm, producing a perishable good, is then

[3]Henceforth, numerical subscripts will denote partial derivatives unless otherwise specified.

[4]Obviously, for a linear demand function with an additive random term, the above approximation becomes exact and our results apply to all probability distributions. Otherwise, for the general random demand function, the results should remain intact for most probability distributions which are compact and concentrated. I am nevertheless aware of the possibility that my results might be sensitive to the choice of particular distributions as James Malcolmson has shown in relation to Martin Weitzman's (1974) problem.

[5]PROOF:

(a) If $q_c < q(p_c,\eta^o)$, then, because $q_1(p,\eta)<0$, there must exist a nonnegative constant $\varepsilon>0$, such that $q_c = q(p_c+\varepsilon,\eta^o)$. Since this function is invertible, we can write $p_c+\varepsilon = p(q_c,\eta^o) \Rightarrow p_c < p(q_c,\eta^o)$. The converse is true since $p_1(q,\eta)<0$.

(b) If $q_c > q(p_c,\eta^o)$, then $q_1(p,\eta)<0$ implies that there exists a nonnegative constant $\delta>0$ such that $q_c = q(p_c-\delta,\eta^o)$. Since this function is invertible, we can write $p_c-\delta = p(q_c,\eta^o) \Rightarrow p_c > p(q_c,\eta^o)$. The converse is trivially true since $p_1(q,\eta)<0$.

[6]It should be noted that even if there is uncertainty in the cost function, the results in this paper are unaltered as long as cost uncertainty is independent of demand uncertainty.

assumed to use the expected profit criterion[7] to evaluate the various behavioral modes which are characterized as follows.

A. *Q Behavior*

In Q behavior, the production or quantity decision is made prior to the revelation of the true demand conditions. The nature of the implicit contract is that the output is assumed to be sold at a price that obtains in the market after the demand conditions are revealed. Denoting $\pi_Q(q)$ as the profit function in Q behavior, the *ex post* profit that obtains after using (3) and (7) is $\pi_Q(q)=p(q,\eta)\cdot q-c(q)$. The formal problem is then to choose an optimal quantity such that expected profit is maximized. Let q^* be optimal. Then q^* satisfies

$$\text{(8)}\quad \operatorname*{Max}_q E[\pi_Q(q)]=E[\pi_Q(q^*)] =E[p(q^*,\eta)]q^*-c(q^*)$$

and by definition of q^*,

$$\text{(9)}\qquad E[\pi_Q(q^*)]\geqslant E[\pi_Q(q)]\quad \forall q\neq q^*$$

B. *P Behavior*

In P behavior, price is set *ex ante* and the nature of the implicit contract is that the firm adjusts its production or quantity to meet the observed demand at the set price. Denoting $\pi_P(p)$ as the profit function relevant for P behavior, the *ex post* profit that obtains after using (2) and (7) is $\pi_P(p)=pq(p,\eta)-c(q(p,\eta))$. The formal problem here is to choose a price such that the expected profit is maximized. Let p^* be optimal. Then p^* satisfies

$$\text{(10)}\quad \operatorname*{Max}_p E[\pi_P(p)]=E[\pi_P(p^*)] =p^*E[q(p^*,\eta)]-E[c(q(p^*,\eta))]$$

and by definition of p^*,

$$\text{(11)}\qquad E[\pi_P(p^*)]\geqslant E[\pi_P(p)]\quad \forall p\neq p^*$$

C. *Q/P Behavior*

In Q/P behavior, first discussed by Edwin Mills, the firm makes its production and price decisions prior to the revelation of the true demand conditions. If actual demand turns out to be below the quantity produced, the firm is left with unsold inventory which is assumed to perish. However, if actual demand exceeds the quantity produced, some customers are left unsatisfied. Denoting $\pi_{Q/P}(q,p)$ as the profit function in Q/P behavior, the *ex post* profit that obtains is

$$\pi_{Q/P}(q,p)=\begin{cases} pq(p,\eta)-c(q) & \text{if } q(p,\eta)\leqslant q\\ pq-c(q) & \text{if } q(p,\eta)\geq q\end{cases}$$

The formal problem is to choose a combination of quantity and price such that expected profit is maximized. Let (q^{**},p^{**}) be the optimal solution. Then (q^{**},p^{**}) satisfies

$$\operatorname*{Max}_{q,p} E[\pi_{Q/P}(q,p)]=E[\pi_{Q/P}(q^{**},p^{**})] =\int_{\underline{M}}^{\eta^{**}} p^{**}q(p^{**},\eta)\,d\Psi(\eta) +\int_{\eta^{**}}^{\overline{M}} p^{**}q^{**}\,d\Psi(\eta)-c(q^{**})$$

where η^{**} satisfies $q(p^{**},\eta^{**})=q^{**}$. The above maximum expected profit can be rewritten alternately as either

$$\text{(12)}\quad E[\pi_{Q/P}(q^{**},p^{**})]=p^{**}q^{**} +p^{**}\cdot I(q^{**},p^{**})-c(q^{**})$$

[7] I have assumed risk neutrality for various reasons. First, the theory of the firm under uncertainty is still unsettled regarding the type of objective function to use. Second, the risk-neutrality assumption is often argued to be more appropriate, because most firms are owned by a large number of shareholders who themselves also hold diversified portfolios. The law of large numbers (see Arrow and Robert Lind) and portfolio diversification suggests that the firm should use expected profit as the appropriate criterion since risk per owner of the firm becomes infinitely small.

or

$$(13)\quad E[\pi_{Q/P}(q^{**},p^{**})]=p^{**}E[q(p^{**},\eta)] + p^{**}S(q^{**},p^{**})-c(q^{**})$$

where

$$(14)\quad I(q^{**},p^{**}) = \int_{\underline{M}}^{\eta^{**}}[q(p^{**},\eta)-q^{**}]d\Psi(\eta)\leqslant 0$$

$$(15)\quad S(q^{**},p^{**}) = \int_{\eta^{**}}^{\overline{M}}[q^{**}-q(p^{**},\eta)]d\Psi(\eta)\leqslant 0$$

Note that $I(q,p)$ and $S(q,p)$ are nonpositive functions for any combination of (q,p). The term $p^{**}I(q^{**},p^{**})$ in (12) is the expected loss in revenue due to unsold inventory which arises whenever demand falls short of production. The term $p^{**}S(q^{**},p^{**})$ in (13), however, is the expected revenue loss due to foregone potential sales whenever demand exceeds production. Thus, the expected sales revenue, which is the sum of the first two right-hand side terms in either (12) or (13), is never greater than either the firm's value of production $p^{**}q^{**}$, or its expected value of demand $p^{**}E[q(p^{**},\eta)]$. Now, given that (q^{**},p^{**}) is by definition optimal,

$$(16)\quad E[\pi_{Q/P}(q^{**},p^{**})] \geqslant E[\pi_{Q/P}(q,p)]\forall(q,p)\neq(q^{**},p^{**})$$

It is important to note that the profit functions, $\pi_Q(\cdot)$, $\pi_P(\cdot)$, and $\pi_{Q/P}(\cdot)$ of the three behavioral modes are different. Consequently, behavioral ranking cannot be simply based on the *ex post* flexibility criterion.

II. Relative Dominance of Behavioral Modes

In comparing the relative dominance of the various behavioral modes, the following propositions are in order for sufficiently small uncertainty.

PROPOSITION 1: *$Q(P)$ behavior is preferred to $P(Q)$ behavior for increasing (decreasing) marginal-cost firms.*

PROOF:

(i) Consider the case of increasing marginal cost which means $c(\cdot)$ is strictly convex. Then, in Q behavior, let q^o be chosen such that

$$(17)\quad q^o=E[q(p^*,\eta)]$$

which, from (6), implies

$$(18)\quad p^*=E[p(q^o,\eta)]$$

Substituting q^o for q^* in $E[\pi_Q(q^*)]$ in (8) and subtracting from it (10), we obtain, after using (17) and (18),

$$(19)\quad E[\pi_Q(q^o)]-E[\pi_P(p^*)] = E[c(q(p^*,\eta))]-c(E[q(p^*,\eta)])>0$$

because of Jensen's inequality since $c(\cdot)$ is strictly convex. Consequently, because of (9), $E[\pi_Q(q^*)>E[\pi_P(p^*)]$ for increasing marginal-cost firms.

(ii) Consider next the case of decreasing marginal cost which implies $c(\cdot)$ is strictly concave. Then in P behavior, let p^o be chosen such that

$$(20)\quad p^o=E[p(q^*,\eta)]$$

which, from (6), implies

$$(21)\quad q^*=E[q(p^o,\eta)]$$

Substituting p^o for p^* in $E[\pi_P(p^*)]$ in (10) and subtracting from it (8) we obtain, after using (20) and (21),

$$(22)\quad E[\pi_P(p^o)]-E[\pi_Q(q^*)] = c(E[q(p^o,\eta)])-E[c(q(p^o,\eta))]>0$$

because of Jensen's inequality since $c(\cdot)$ is strictly concave. Consequently, because of (11), $E[\pi_P(p^*)]>E[\pi_Q(q^*)]$ for decreasing marginal-cost firms.

From the above proof, it is easily deduced that under constant marginal costs and/or in the absence of demand uncertainty, the firm would be indifferent between Q and P

behavior. This highlights the fact that the issue of choice of behavioral modes is irrelevant in the absence of uncertainty but is not so when uncertainty exists.

The results in Proposition 1 can be interpreted as follows. Given identical expected revenues in both behavioral modes, the differences in their expected profits depend on their differences in expected costs. In P behavior, the firm adjusts its production, *ex post*, according to the observed demand, and hence production is viewed as random before price is set. Contrast this with Q behavior where production is fixed. But we know that, given the same expected production in both cases, the expected cost of a random production strategy exceeds (falls below) that of a fixed production strategy under increasing (decreasing) marginal-cost conditions. Hence it is not surprising to find the fixed production (Q behavior) being preferred to the random production (P behavior) strategy in the case of increasing marginal cost and the opposite in the case of decreasing marginal cost.

PROPOSITION 2: *Q behavior is preferred to Q/P behavior for either increasing or decreasing marginal-cost firms.*

PROOF:

In Q/P behavior, the solution (q^{**},p^{**}) is by definition, optimal. However, depending on the nature of demand and cost functions, the relationship between q^{**} and p^{**} could take either of the following forms:

$$E[q(p^{**},\eta)] \geqslant q^{**}; \quad E[q(p^{**},\eta)] \leqslant q^{**}$$

(i) Consider the case where $E[q(p^{**},\eta)] \geqslant q^{**}$. This, from (6), implies

$$p^{**} \leqslant E[p(q^{**},\eta)] \tag{23}$$

The maximum expected profit in Q/P behavior expressed in (12) is rewritten as

$$E[\pi_{Q/P}(q^{**},p^{**})] = p^{**}q^{**} + p^{**}I(q^{**},p^{**}) - c(q^{**}) \tag{24}$$

In Q behavior, let q^{**} be arbitrarily chosen as a solution. Substituting q^{**} for q^* in $E[\pi_Q(q^*)]$ in (8), the expected profit from this choice of q^{**} under Q behavior is

$$E[\pi_Q(q^{**})] = E[p(q^{**},\eta)]q^{**} - c(q^{**}) \tag{25}$$

Subtracting (24) from (25),

$$\begin{aligned} E[\pi_Q(q^{**})] - E[\pi_{Q/P}(q^{**},p^{**})] &= [E[p(q^{**},\eta)] - p^{**}]q^{**} \\ &\quad - p^{**}I(q^{**},p^{**}) \geqslant 0 \end{aligned} \tag{26}$$

because of the inequalities (23) and (14). Consequently, because of (9),

$$E[\pi_Q(q^*)] \geqslant E[\pi_{Q/P}(q^{**},p^{**})]$$

(ii) Consider next the case where $E[q(p^{**},\eta)] \leqslant q^{**}$. The maximum expected profit under Q/P behavior expressed in (13) is rewritten as

$$E[\pi_{Q/P}(q^{**},p^{**})] = p^{**}E[q(p^{**},\eta)] + p^{**}S(q^{**},p^{**}) - c(q^{**}) \tag{27}$$

In Q behavior, however, let q' be arbitrarily chosen as a solution such that

$$q' = E[q(p^{**},\eta)] \tag{28}$$

which by (6) implies

$$p^{**} = E[p(q',\eta)] \tag{29}$$

Also, note that by assumption, $E[q(p^{**},\eta)] \leqslant q^{**}$, so that from (28),

$$q' \leqslant q^{**} \tag{30}$$

Substituting q' for q^* in $E[\pi_Q(q^*)]$ in (8), the expected profit in Q behavior from the arbitrary choice of q' is

$$E[\pi_Q(q')] = E[p(q',\eta)]q' - c(q') \tag{31}$$

Subtracting (27) from (31), and using (28) and (29), we obtain

$$\begin{aligned} E[\pi_Q(q')] - E[\pi_{Q/P}(q^{**},p^{**}) \\ = -p^{**}S(q^{**},p^{**}) + [c(q^{**}) - c(q')] \geqslant 0 \end{aligned} \tag{32}$$

because of inequalities (15) and (30) and also the fact that $c'(\cdot)>0$. Consequently, because of (9),

$$E[\pi_Q(q^*)] \geqslant E[\pi_{Q/P}(q^{**},p^{**})]$$

The superiority of Q over Q/P behavior, as shown in the proof, mainly arises from the disadvantage of Q/P behavior in having to suffer losses in either unsold inventory or foregone potential sales. This negative feature does not occur in Q behavior in which the firm is able to sell all its output *ex post* at some price given by the market.

In the case of relative dominance between P and Q/P behavior, we establish the following results.

PROPOSITION 3: *P behavior is preferred to Q behavior which in turn is preferred to Q/P behavior for decreasing marginal-cost firms.*

PROOF:

This immediately follows from Propositions 1 and 2.

PROPOSITION 4: *P behavior is not always preferred to Q/P behavior for increasing marginal-cost firms.*

PROOF:

In Q/P behavior, let an arbitrary solution (q^o,p^*) be chosen such that the relationship between q^o and p^* is the following:

$$q^o = E[q(p^*,\eta)] \tag{33}$$

Substituting (q^o,p^*) for (q^{**},p^{**}) in (12), and then making use of (33), the expected profit in Q/P behavior,

$$E[\pi_{Q/P}(q^o,p^*)] = p^*E[q(p^*,\eta)] + p^*I(q^o,p^*) - c(E[q(p^*,\eta)]) \tag{34}$$

Subtracting (10) from (34),

$$E[\pi_{Q/P}(q^o,p^*)] - E[\pi_P(p^*)] = p^*I(q^o,p^*) + E[c(q(p^*,\eta))] - c(E[q(p^*,\eta)]) \tag{35}$$

In (35), the first right-hand side term is nonpositive. The second right-hand side term, however, is positive because of Jensen's inequality since $c(\cdot)$ is strictly convex for increasing marginal cost. Nonetheless, because $I(q^o,p^*)$ depends strictly on the nature of demand conditions and these are totally unrelated to the cost conditions, we can always construct an arbitrary example by independently choosing some demand and cost conditions, such that (35) is nonnegative.

As an illustration, consider a simple example of a linear demand function $q(p,\eta) = a - dp + \eta$ with $E[\eta]=0$. Then, using (33), the term

$$p^*I(q^o,p^*) = p^*\int_{\underline{M}}^{0} \eta \, d\Psi(\eta)$$

$$\equiv p^*E[\eta|\eta \leqslant 0] \leqslant 0$$

Now, consider a quadratic cost function, $c(q) = c_0 + c_1q + (1/2)c_{11}q^2$ where $c_{11}>0$ is the slope of the marginal-cost function. Then the expression

$$E[c(q(p^*,\eta))] - c(E[q(p^*,\eta)]) = (1/2)c_{11}\sigma^2 \geqslant 0$$

where σ^2 is the variance of the random variable η. Using this construction, the expression (35) becomes

$$E[\pi_{Q/P}(q^o,p^*)] - E[\pi_P(p^*)] = p^*E[\eta|\eta \leqslant 0] + (1/2)c_{11}\sigma^2$$

which can be nonnegative if

$$c_{11} \geqslant -\frac{2p^*E[\eta|\eta \leqslant 0]}{\sigma^2}$$

Thus, Q/P can dominate P behavior for sufficiently increasing marginal-cost firms.

The above result confirms my earlier contention that because the profit functions differ among behavioral modes, the use of *ex*

post flexibility criterion, which may lead one to infer that P behavior is always preferred to Q/P behavior, is misleading. In observing (35), it is clear why there is ambiguity in preference between P and Q/P behavior in the case of increasing marginal cost. The first right-hand side term of (35) reflects the advantage of P over Q/P behavior in that the former strategy, unlike the latter, need not incur any loss due to unsold inventory. The second right-hand side term of (35), however, reflects the advantage of Q/P over P behavior because, for the same expected production, the expected cost of a random production (P behavior) exceeds that of a fixed production (Q/P behavior) strategy in the case of increasing marginal cost. These two counteracting effects make the preference between P and Q/P behavior ambiguous for increasing marginal-cost firms.

III. Conclusion

In summary, it is shown that for decreasing marginal-cost firms, P behavior is preferred to Q behavior which in turn is preferred to Q/P behavior. In the case of increasing marginal cost, Q behavior is preferred to P and also Q/P behavior, but the preference between the latter two is ambiguous, depending on the nature of demand and cost conditions. The interpretation of these results can be summarized as follows.

In the case of increasing marginal cost, Q is preferred to P as well as Q/P behavior for two separate reasons. Mainly because of cost consideration Q is preferred to P behavior since increasing marginal cost, given the same expected production tends to favor a fixed production (Q behavior) over a random production (P behavior) strategy. Whereas, mainly due to the randomness in demand, Q is preferred to Q/P behavior because of the relative disadvantage of the latter strategy in having to absorb the risk of unsold inventory. Since the cost and random demand conditions are unrelated, it is not surprising that the relative dominance between P and Q/P behavior is ambiguous. However, the larger the slope of the marginal cost function, *ceteris paribus*, the less attractive would be P behavior, which involves a random production strategy, relative to Q/P behavior whose production is fixed.

In the case of decreasing marginal cost, the above ambiguity in preference between P and Q/P behavior disappears. First, decreasing marginal cost, given the same expected production, favors a random production (P behavior) over a fixed production (Q behavior) strategy. Second, Q in turn is preferred to Q/P behavior because of the risk of unsold inventory which is a characteristic of the latter behavior.

Notice that regardless of the nature of the cost function, Q is always preferred to Q/P behavior. The immediate implication of this result is that if both Q and Q/P behavior are feasible, we should expect Q/P behavior to be nonexistent which is contrary to casual empiricism where firms are often observed to set both price and quantity. This raises an interesting issue of whether or not Q behavior is always a feasible choice. Recall that implicit in Q behavior is that the firm is able to sell all its output at some price that obtains in the market. The exact nature of the market arrangement that allows for the existence of Q behavior is unclear, since it is not specified in any detail in our model. I conjecture, however, that the feasibility of Q behavior may require the existence of some marketing agency that operates to purchase all of the firm's quantity at some price. Such a marketing agency may be in the form of intermediate trading agents (for example, retailers). The very domination of Q over Q/P behavior suggests that firms should be willing, in order to operate under Q behavior, to offer price discounts for bulk sales to intermediate trading agents rather than having to conduct sales directly under Q/P behavior where it directly assumes the risk of unsold inventory. To the extent that the price discount offered by the firm is sufficient to cover the normal profits of intermediate trading agents, such risk is effectively shifted from the former to the latter under Q behavior. Otherwise, if the price discount is insufficient to attract entry of intermediate trading agents, then the firm itself

may have to be contented to operate under Q/P behavior. This conjecture needs to be further explored in an explicit model of optimal risk sharing.

REFERENCES

K. J. Arrow, "Toward a Theory of Price Adjustment," in Moses Abramovitz, ed., *The Allocation of Economic Resources*, Stanford 1959, 41–51.

______ **and R. A. Lind,** "Uncertainty and the Evaluation of Public Investment Decisions," *Amer. Econ. Rev.*, June 1970, *60*, 364–78.

D. Baron, "Demand Uncertainty in Imperfect Competition," *Int. Econ. Rev.*, June 1971, *12*, 196–208.

H. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

C. Lim, "Uncertainty and Implicit Contracts," unpublished doctoral dissertation, Queen's Univ. 1978.

J. M. Malcolmson, "Prices vs. Quantities: A Critical Note on the Use of Approximations," *Rev. Econ. Stud.*, Feb. 1978, *45*, 203–07.

E. S. Mills, "Uncertainty and Price Theory," *Quart. J. Econ.*, Feb. 1959, *73*, 116–29.

P. A. Samuelson, "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variance, and Higher Moments," *Rev. Econ. Stud.*, Oct. 1970, *37*, 537–42.

G. J. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 213–25.

M. L. Weitzman, "Prices vs. Quantities," *Rev. Econ. Stud.*, Oct. 1974, *41*, 479–91.

______, "Reply to 'Prices vs. Quantities: A Critical Note on the Use of Approximations' by James M. Malcolmson," *Rev. Econ. Stud.*, Feb. 1978, *45*, 209–10.

Many, Few, One: Social Harmony and the Shrunken Choice Set

By DAVID FRIEDMAN*

An individual, to choose, must know what he wants to achieve and how to achieve it. Positive economics attempts to answer the second question by establishing propositions about what consequences will follow from particular acts. For an individual who already knows what ends he wishes to achieve, that is sufficient. But many, perhaps most, of the questions economics is asked to deal with involve choices that affect many people at once. It appears that positive economics cannot generate any recommendations concerning such choices unless it is combined with some way either of aggregating preferences or of choosing among them.

My thesis is that this conclusion is false, that the methods of positive economics can, and to some small degree do, yield recommendations for choices to be made jointly by a large number of people, recommendations which do not depend on some normative scheme for going "behind" their separate preferences or combining them in some "justly balanced" fashion.

Since each individual acts in terms of his values, the fact that something is a Pareto improvement induces everyone to agree that it is desirable and act accordingly—even though they may all be wrong "in the eyes of God." In this sense a positive proposition may generate widespread, even unanimous, assent to a normative proposition. But although a change may be a Pareto improvement, it seems unlikely that out of numerous possible changes the same one will be preferred by everyone. This is the familiar problem of choosing among points on the Pareto frontier.

If the outcomes advocated by people of varying political viewpoints were all possible there might be no way to reach agreement without first resolving normative differences. But one of the functions of positive economics is, by finding the consequences of acts and institutions, to determine what outcomes are possible. If the number of possible outcomes turns out to be much smaller than the number that people imagine it may be much easier than we would at first expect to achieve agreement about which is most desirable.[1] To put it differently, it is a positive question whether the social opportunity set is like that shown in Figure 1a or in Figure 1b. Even if we cannot produce unanimity by reducing the frontier to a single point, a demonstration that the choice set was similar to Figure 1c would drastically reduce the importance of achieving one point on the frontier rather than another, and hence the resources spent trying to do so.

It is an interesting empirical observation that people with differing political views almost always disagree widely about positive propositions.[2] The socialist and the libertarian disagree about the relative justice of capitalist and socialist institutions. But they also disagree, often even more, on how these institutions do or will work; the capitalist society that the socialist attacks barely has a family resemblance to the capitalist society that the libertarian defends. And I have yet to meet the socialist who is willing to defend the society that

*Virginia Polytechnic Institute and State University. I would like to express my appreciation to the Liberty Fund for sponsoring the conference for which this paper was originally written, and to thank James Buchanan for first bringing Schelling points to my attention.

[1]It is also possible that the reduction of alternatives may eliminate possible compromises and so make agreement more difficult; this point is due to H. G. Brennan.

[2]Some of the most striking examples concern historical propositions such as the factual disagreements discussed in Friedrich Hayek and Robert Wasson.

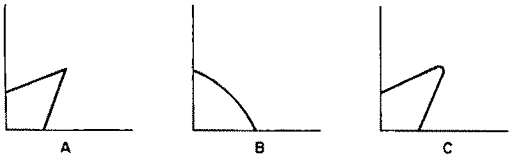


FIGURE 1

Hayek believes socialism would produce.[3] Even in arguments on abstract justice the socialist is reluctant to pronounce judgments on the libertarian's hypothetical situations (in which the capital of the capitalist consists of a field hacked from the jungle by the sweat of his brow while the lazy worker looked on) and the libertarian equally reluctant to accept the socialist's hypotheticals as examples of the working of a "truly capitalist" society.[4] All of which suggests that political disagreements, disagreements about how men should interact and to what ends, although they may involve normative disagreements involve positive ones as well and at perhaps as fundamental a level. If the positive questions were resolved, if the one disputant were convinced that the real alternatives were those proposed by the other or if both were convinced of a common set of alternatives, both, or all, might agree, not about what sort of society they wanted, but about what society was the best they could achieve.

To some small extent this has already happened. The desirability of the market mechanism as an *allocational* device is widely accepted by people of diverse political and ethical positions, from anarchists to communists.[5] This agreement is not based on unanimity concerning normative propositions, but on agreement about positive propositions that describe how the market and its alternatives function. The agreement is far from unanimous; there are many communists, socialists, dirigiste liberals, and others, who do not accept the advantages of the market. But their disagreement with their "coreligionists" who do accept it is over positive, not normative, questions.

The agreement between the market socialist and the libertarian breaks down over distribution. Here again we have agreement on the desirability of achieving the Pareto frontier combined with disagreement as to where on that frontier we should be. But in assuming that we are free to choose among many points on the Pareto frontier which differ substantially in terms of distribution, but only slightly in what is available to be distributed, we implicitly assume that transfers are both possible and inexpensive. That is a positive proposition. If one could demonstrate that the distributional alternatives are far more limited than is usually assumed, one might again generate agreement on which alternative is best without first requiring agreement on the meaning of good. The remainder of this paper is devoted to an exploration of two approaches to narrowing the range of possible alternatives. The first attempts to explore certain general characteristics of mechanisms for transferring income, characteristics which suggest that such mechanisms may be too expensive to be attractive even to those who approve of the direction of the transfer. The second combines elements of the first with certain observations about bargaining to suggest that we may be limited to a choice between only two distributional alternatives —equality and the status quo.

Before continuing, I should warn the reader that these are areas where relatively little work has been done and he or she may therefore find that the remainder of the paper contains a ha'penny of argument for a pound of speculation. But the purpose of this paper is not to establish propositions in positive economics which will compel universal assent and so end forever all political strife. My thesis is only that it is possible for economics to establish propositions which generate agreement on choices without first

[3]George Orwell's response to *The Road to Serfdom* in Sonia Orwell is an example of the reaction of a socialist who feared that Hayek's description might be correct.

[4]A good example in fiction is the contrast between the capitalist portrayed by George Bernard Shaw and by Ayn Rand. Compare, for instance, the preface to *Major Barbara*, especially pp. 34–35, with the portrayal of Hank Reardon in *Atlas Shrugged*.

[5]For anarchist examples, see my book and Murray Rothbard. Socialist examples are Abba Lerner, James Meade (1948, 1965), and Ota Sik.

requiring agreement on ends. The purpose of the following sections is to suggest where some such propositions might be found—not to find and prove them.

I. Krueger's Law and the Marginal Impossibility of Transfers[6]

Let us suppose that the government wishes to subsidize a particular class of people. Suppose further that it is possible, at some cost, to become a member of the class. Obvious examples are the unemployed, college students, and unmarried mothers. Since people enter the class until the marginal cost of doing so to the last entrant is equal to the subsidy he receives, there is at the margin no net benefit to the recipient. A net benefit to recipients exists only because some people are not at the margin—they are able to enter the class at a cost lower than the cost to the marginal individual. While the cost to the government, which we would commonly describe as the amount of the subsidy, is the total sum spent, the benefit to the recipients is only the consumer's (or, if you prefer, producer's) surplus on that expenditure.[7] Expenditures on the unemployed represent a transfer to them only in the sense in which purchases of paper clips represent a transfer to the manufacturers of paper clips.

The argument can be carried further. Suppose there are many nonmarginal members of the group and they receive a substantial benefit from the program. So far we have considered the membership of the group as variable and the existence of the program as given. Now reverse the assumptions. In order to get the subsidy passed, the potential beneficiaries are willing, if necessary, to spend any amount up to the total benefit they expect—that is to say, the consumer's surplus they expect to receive. If, for those in a position to decide on the subsidy (legislators, for example), the choice is between subsidizing or not subsidizing that group, then there is a bilateral monopoly bargaining situation between the legislators and the beneficiaries; the exact "price" for which the legislation will be sold is indeterminate, but will presumably be a substantial fraction of its value to the recipients. If, as seems more probable, the legislators are faced, because of limited resources, with the choice of which one of various groups to subsidize, the competition among those groups should allow the legislators to charge the recipients the full value to them of the subsidy. In this case the "beneficiaries" of the subsidy end with no net benefit at all—part of the subsidy has been spent by them individually to pay the costs of entering the group and so qualifying themselves to receive the subsidy, and the rest has been spent collectively on inducing the legislators to provide the subsidy. Just as in the familiar case of an ordinary competitive market, all profits have been competed away. If, as a third possibility, the legislators compete among themselves for the privilege of "selling" the subsidy, the beneficiaries should be able to retain most or all of the surplus (net of any cost to the politicians of passing the subsidy).[8]

In what sense is this a normal and in what sense a perverse outcome? It is normal if we think of the transaction as equivalent to the purchase of paper clips. In that case too, competition to enter the group—to sell the government paper clips—guarantees that the marginal member receives no net profit. And in that case too, a monopsonistic purchaser (the government) should, under favorable bargaining circumstances, be able to appropriate any producer's surplus.

But the purpose of buying paper clips is to provide paper clips not income for those who manufacture them. The "purpose" of a government transfer payment is to make the recipients better off. For paper clip manufacturers to compete away their profits is desirable; for the recipients of a transfer to do so destroys the entire point of the transfer.

[6]Earlier statements of the argument are Gordon Tullock (1967, 1971) and my book, pp. 207–12.

[7]This is, of course, only another way of looking at the argument usually summarized in taxation under the rubric of "excess burden" and elsewhere as "disincentives."

[8]Presumably a two-party system, by guaranteeing that there exists a majority coalition among the legislators, discourages such competition and so raises the rents received by the legislators. This may be one reason such systems exist.

Does the process at least result in a net benefit to the legislators? Not necessarily. It raises their income, but one consequence of that is to encourage other people to compete for their positions, raising the total amount spent on campaign advertising and similar expenses. There may be some producer's surplus to the politician—a nonmarginal politician may be able to get elected even though he spends less than the total amount he will receive if elected times the probability of winning. Nonmarginal politicians are presumably of two sorts. Some have innate talents on which they receive rent. Others are nonmarginal because they are incumbents. They receive quasi rents purchased in competition with other equally talented individuals when they first ran for the seat.

It appears from this argument that if \$10 million is spent on transfers, the recipients will be unaffected and the society as a whole will be poorer by almost exactly \$10 million. There are several reasons why this might not occur. If the recipients are an unorganized group they may be unable to put together, as a bribe to the legislators, the full amount of the benefit they expect to receive; in collecting the money they are faced with the usual public good problem of organizing a small group.[9] After offering the largest payment they can collect, they may end up with some net benefit—or the legislators may provide the subsidy to some other and better organized group instead.[10]

Alternatively, the group may be especially qualified to receive the subsidy. If one motivation for the subsidy is public desire to help that particular group, the legislators are not entirely free to give the subsidy to some other group instead—doing so may cost them votes. In this case the situation reverts to a bilateral monopoly and the beneficiaries may hope, by successful bargaining with the legislators, to retain part of the benefit for themselves.

I do not claim that this line of analysis conclusively proves anything about the consequences of governmental redistribution; the careful (and especially the hostile) reader will have noticed that I have implicitly introduced a set of assumptions about the functioning of the political market and the motivations of its participants with which many might quarrel. But those assumptions are positive, not normative, and both the assumptions and the conclusions they lead to are testable.

What then are the conclusions which this argument suggests, and which further work along the same lines might prove? They are that the existence of a mechanism for redistribution results in very large net costs and very little actual redistribution. If this conclusion is correct, it might lead even the most ardent proponents of such mechanisms to conclude that they are undesirable.

Moving, for a moment, from the abstract to the at least partly concrete, it is worth noting that Ann Krueger, in an article based on arguments of this sort, concluded that policies which appeared to give special benefits to certain businessmen in India and Turkey actually had the effect of wasting on the order of 10 percent of the national income of each country.

II. Bargaining Costs, Schelling Points, and the Status Quo

I have argued that the existence of mechanisms for transferring wealth results in large and unproductive expenditures motivated by the desires of potential recipients to receive transfers. But mechanisms for transferring wealth always exist. Some are formally legal—form a majority, elect politicians, appoint judges, then do what you want. Some are formally illegal—coup, revolution, theft. The arguments made earlier suggest that the existence of such possibilities should be very costly. If any group that can form a majority coalition can proceed to transfer the resources of the minority to its members, the attempt to form such coalitions, and prevent other people from forming them, should be very nearly a full-time occupation.

[9]See Mancur Olson.

[10]This argument suggests that concentrated (or unionized) industries will be the most likely to receive tariff protection, subsidies, etc., but will get the least net return from those benefits they receive since they will have bought them at a high price.

Yet somehow we survive. Jimmy Carter, despite majorities in both houses, is not expected to deprive Gerald Ford of his pension and expropriate his supporters. Although the government has almost unlimited power to redistribute income, the actual distribution of income seems to remain relatively stable for decades, at least in comparison with the wild fluctuations one would expect to result from an unlimited stakes majority-rule bargaining game.[11] One can attempt to explain why this is so by starting either with abstract considerations of the bargaining process or with an examination of implicit features of our "social contract" as reflected in the beliefs and prejudices of the society's members. I will do both.

Imagine a group of people engaged in a game to determine how they will divide up a pot containing a certain amount of money. The players' information about the game is sufficiently imperfect that they have only approximate and inconsistent estimates of how much each can expect to end up with. At any time the players can stop the game and divide up the pot, provided that they can agree on a division. Assume further that the process of playing, and of bargaining about the division, is itself costly; every five minutes the amount in the pot declines by a dollar.

The players could all be better off if they stopped playing and agreed on some division of the pot within the first five minutes. One player will therefore propose some division, reflecting his estimate of the expected returns of the different players, and probably biased in his own favor. The next player, suspecting the bias and having a different and perhaps higher estimate of his own expected return, rejects that proposal and substitutes one of his own, which is in turn rejected by the maker of the first offer and the other players on similar grounds. It is in everyone's interest to reach some agreement but any particular agreement is more in the interest of some than of others. The bargaining (interspersed with intervals of play, in the hope that as the game continues the real returns that the players can expect will become clearer) may continue beyond five minutes and may even continue until the game is played out or the pot empty.

Let us now add one more element. Before the game begins, the players are handed a "suggested division." The suggested division corresponds very approximately to the players' real playing strength—that is to say to the proportional returns they can expect if the game is played out—but no more closely than do the estimates of the individual players. It seems likely ("seems" only, since no satisfactory theory of bargaining exists) that this addition, which in no way affects the rules of the game, will change the outcome; that the same set of players who would have bargained interminably over each other's suggested divisions will accept the suggested division in the first five minutes.

Why? Without the suggested division, each player is faced with the choice among many alternative suggestions, and has the opportunity of adding his own suggestion and having it considered on the same basis as all the others. Once the suggested division exists, the choice is seen not as between "this division, and that division, and that other division, and making a new suggestion, and..." but between "accepting the suggested division or bargaining some more." The player has no reason to expect further bargaining to raise his share of the pot and he knows that the pot is shrinking. So he accepts the suggested division instead.

This is an example of a phenomenon described by Thomas Schelling (ch. 3) and called (by others) the Schelling point. It seems, empirically and intuitively, that bargaining tends to converge on outcomes perceived as unique. When bargaining must choose among a multitude, or from a continuum, of possible outcomes, each more attractive to some and less to others, it is difficult to even reach a resolution—each

[11]Herman Miller concludes that the shape of the income distribution has been very stable over time. This does not, of course, prove that the people occupying various positions in the distribution have remained the same. Morton Paglin uses a different definition of equality to argue that some changes in recent decades have been somewhat larger than generally believed.

offer by one party calls forth a slightly different offer by the other. If the bargaining process is costly, everyone may prefer to accept an outcome which, because of its uniqueness, is perceived as presenting the bargainers with the choice "this or continued bargaining," not "this, or that, or...." Obvious examples of Schelling points are splitting a profit fifty-fifty or letting a disputed boundary run along some convenient stream. It is worth noting that these Schelling points are not imposed by some higher authority, but rather given by the nature of the situation. More precisely, the existence of a Schelling point is a "fact" about how the bargainers perceive the structure of the alternatives they face.

One could discuss the logic of Schelling points at much greater length. Instead, I suggest that those readers who are not convinced think about why, when a sum must be somehow divided between two men, an even split seems the "natural" solution.

Having now discussed the abstract game and its "solution," let us return to the real political (or revolutionary) game. We could choose to play on the assumption that every temporary winner takes all he can get. The game would be extraordinarily expensive; the players would choose to expend large resources trying to win. It is therefore in our joint interest to agree on some initial distribution in advance, a distribution based loosely on our guesses about what the final outcome of the game would be, and to then play the political game under much more restrictive rules, rules which guarantee that that distribution is substantially maintained and thereby drastically lower the stakes we are playing for. In choosing the initial distribution we face the same problem as the players of my imaginary game. Each player, each faction, will propose his own division; the process of bargaining, with threats, counteroffers, coalitions, betrayals, rapidly becomes the full-scale political game itself, with all its costs. To rescue us from this situation we require a Schelling point—a particular distribution of wealth, corresponding at least very roughly with our guesses about the final outcome of the game, that is seen not as the proposal of one faction but as an alternative "offered by nature" to the nightmare of continued dispute. There are, I think, only two candidates. One is the status quo distribution, the other is equality.[12]

Earlier I suggested that one could try to explain the paradox of civil society by examining the beliefs and prejudices of its members. Having reached my conclusion through an entirely abstract argument, I will now try to show that that conclusion accurately describes the way our society has resolved the problem—insofar as it is resolved.

Imagine that, during the oil shortage, someone suggests price control in order to hold the price of gasoline ten cents below the level it would otherwise rise to, and justifies it by showing that a rise in the price of gasoline will injure the poor. Suppose it were argued on the other side that such price control would result in a shortage. I think that most Americans would have felt that the prospect of a shortage was a serious argument against price control and the need to protect the poor a serious argument in favor. Both arguments received widespread consideration.[13]

Suppose now that before the oil shortage someone had proposed price controlling gasoline at a price ten cents below its then

[12]Neither equality nor the status quo is precisely defined, so that even after agreement on one of the Schelling points there may still remain an (expensive) bargaining game "within"it. One of the attractions of a system of property rights is that it provides a definition (not necessarily the only possible one) of a "moving status quo," a set of rules clearly defining how, in a changing world, gains and losses are to be allocated. In order for this to happen, the set of property rights must be perceived, if not as just, at least as somehow special, unique. If perceptions change in such a way that for the owner to receive the benefit of a rise in his house's value is seen as "no more natural than" for anyone else to receive it, property rights cease to perform this function and every change sets off a new distributional struggle. It must be remembered that the existence of a Schelling point is a subjective fact, a fact about how people perceive the alternatives, and that subjective facts can change. One might argue that in the United States in recent decades this Schelling point has been gradually dissolving, as witness, for example, the recent unpopularity of "windfall gains."

[13]J. W. Anderson's article in the November 28, 1973 *Washington Post* explores the difficulties mentioned in fn. 12 as applied to gas rationing.

market price in order to force its price down and so *benefit* the poor. Again the argument against would be the prospect of a shortage. This time very few people would take the argument for price control seriously, and in fact no such argument was widely considered.

Apparently people think it reasonable to bear a substantial cost (a shortage) in order to protect the poor against an injury of a certain amount, but not reasonable to bear a similar cost to give the poor a positive benefit of the same size. This would make sense if people believed that the existing distribution of income were in some way particularly just or desirable so that anything which disturbed it in either direction was bad. But, as a matter of casual observation, nobody does believe that.

My interpretation is that these views reflect an implicit acceptance of the special position of the status quo. Government action to substantially alter the status quo is perceived as of dubious legitimacy; government action to preserve it is not.[14]

It may be objected that in fact government spends most of its time transferring money from one group to another and so alters the status quo. My reply is that while government, in the process of doing things which are argued for on other grounds, may help some people and hurt others, redistribution for the sake of redistribution, in our society and I suspect in most others, is widely perceived as illegitimate.

The distinction here may be a subtle one. It is possible in our society to argue for a government program to help the poor. But the argument is not that the poor, being part of the winning coalition, should benefit at the expense of others. The argument is that by helping the poor we can make everyone better off, that helping the poor is not merely a means to make the poor happier but a means to reduce crime, make us all feel less guilty, make the cities livable, or whatever. What may, from the standpoint of wealth, be a (small) redistribution is defended as, from the standpoint of utility, a Pareto improvement.

The purpose of this paper, as I have said before, is not to prove positive propositions about the workings of political or economic markets. The arguments of this section do not lead to a rigorous conclusion about the degree to which government will or will not redistribute. But they suggest that in trying to analyze the vast bargaining problem in which we are all participants we may find our alternatives far more restricted than we would expect. We may find that the cost of having a society in which the question of how wealth is to be distributed is left open is that there is no wealth left to be distributed. That proposition, whether true or false, is positive, not normative; if it turns out to be true, we may find that the distributional question, like the allocational question, has a single answer which will command widespread support.

[14]See, for example, the *New York Times* editorial of November 15, 1973, p. 44, and the *Washington Post* editorial of November 30, 1973, p. A30.

REFERENCES

James M. Buchanan, *The Limits of Liberty: Between Anarchy and Leviathan*, Chicago; London 1975.

Gerard Debreu, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New Haven; London 1959.

David Friedman, *The Machinery of Freedom: Guide to a Radical Capitalism*, New York 1973.

Friedrich A. Hayek, *Capitalism and the Historians*, Chicago 1954.

A. Krueger, "The Political Economy of the Rent Seeking Society," *Amer. Econ. Rev.*, June 1974, *64*, 291–303.

Abba Lerner, *The Economics of Control*, New York 1947.

James E. Meade, *Planning and the Price Mechanism—The Liberal Socialist Solution*, London 1948.

______, *Efficiency, Equality and the Ownership of Property*, Cambridge 1965.

Herman P. Miller, *Income Distribution in the United States*, A 1960 Census Monograph, Washington 1966.

Mancur Olson, Jr., *The Logic of Collection Action: Public Goods and the Theory of*

Groups, Cambridge 1965.

Sonia Orwell, *The Collected Essays, Journalism and Letters of George Orwell*, Vol. 3: *As I Please*, New York 1968.

M. Paglin, "The Measurement and Trend of Inequality: A Basic Revision," *Amer. Econ. Rev.*, Sept. 1975, *65*, 598–609.

Ayn Rand, *Atlas Shrugged*, New York 1957.

Murray Rothbard, *For A New Liberty*, New York 1973.

Thomas C. Schelling, *The Strategy of Conflict*, Oxford 1960.

George Bernard Shaw, *Major Barbara*, Baltimore 1959.

Ota Sik, *Plan and Market Under Socialism*, Prague 1967.

Gordon Tullock, *Explorations in the Theory of Anarchy*, Blacksburg 1972.

______, "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Econ. J.*, No. 3, 1967, *5*, 224–32.

______, "The Costs of Transfers," *Kyklos*, No. 4, 1971, *24*, 629–43.

Robert G. Wasson, *The Hall Carbine Affair*, New York 1948.

Washington Post, Nov. 28, 1973; Nov. 30, 1973, p. A30.

New York Times, Nov. 15, 1973.

Distributional Neutrality and Optimal Commodity Taxation: Comment

By Leslie Young*

The pioneering papers on optimal commodity taxation have concentrated on the case of an economy with a single consumer or many identical consumers. In a recent paper in this *Review*, David Wildasin discussed the validity of the resulting rules in the many-consumer case under two restrictions that can be imposed on the social welfare function, which constitute alternative formalizations of the idea of "distributional neutrality." In this note I show that one of his central propositions is false. This leads to a modification and simplification of his conclusions on the role of "neutrality" assumptions in optimal commodity taxation.

I. Wildasin's Model and Results

I shall use the model and notation of Wildasin's paper, to which the reader is also referred for a discussion of the literature. There are H households, indexed by h, and $n+1$ goods, indexed by i. The government production vector y^g is fixed throughout. The private production vector y^f satisfies the technical constraint:

$$\phi^f y^f \leqslant 0$$

where ϕ^f is a fixed, strictly positive vector. The term x^h denotes the net consumption vector of household h, whose preferences are represented by a strictly quasi-concave, differentiable utility function $u^h(x^h)$. The term q is the vector of consumer prices, p the vector of producer prices, and $t=q-p$ the vector of commodity taxes. The indirect utility function of household h is $v^h(q,I^h)$ where I^h is its income.

The government maximizes the social welfare function $W(v^1,\ldots,v^H)$ by choosing consumer and producer prices. As Wildasin remarks, this is equivalent to the problem:

$$\text{(1)} \qquad \max_{q,y^f}\left[\,W(v^1,\ldots,v^H):y^f+y^g=X,\phi^f y^f\leqslant 0\,\right]$$

where $X=\sum_h x^h$. The usual analysis yields the first-order conditions

$$\text{(2)} \qquad \sum_h \frac{W_h v_I^h}{\rho_0}x_k^h=\frac{\partial(tX)}{\partial t_k} \qquad k=0,\ldots,n$$

where ρ_0 is the Lagrange multiplier associated with the constraint $y_0^f+y_0^g=X_0$ on the numeraire commodity 0.

In the case of a single consumer (2) becomes

$$\text{(3)} \qquad \frac{v_I}{\rho_0}=\frac{1}{X_k}\frac{\partial(tX)}{\partial t_k} \qquad k=0,\ldots,n$$

From (3) the Slutsky relations yield

$$\text{(4)} \qquad \frac{v_I}{\rho_0}-1+t\frac{\partial X}{\partial I}=\sum_{i=0}^{n}\frac{t_i}{X_k}\left.\frac{\partial X_k}{\partial t_i}\right|_{\bar u}$$

Equation (3) is the rule of Marginal Revenue Proportionality: For every commodity k, the ratio of the marginal tax revenues associated with a marginal increase in the tax rate on k to the quantity of k purchased is the same.

Equation (4) is the Ramsey Rule: For all commodities, the percentage reduction in demand (along the compensated demand curve) due to taxation is (approximately) the same.

For an economy with many consumers, Wildasin introduces two assumptions

Simple Neutrality:

$$W_h v_I^h\equiv\mu;\ h=1,\ldots,H \text{ for fixed } \mu$$

* Senior lecturer in economics, University of Canterbury, New Zealand.

Extended Neutrality:

$$\frac{W_h v_I^h}{\rho_0} + \frac{\partial(tx^h)}{\partial I^h} \equiv 1-\theta, \qquad h=1,\ldots,H \text{ for fixed } \theta$$

Simple neutrality requires that the marginal social utility of consumption be the same for every household. Extended neutrality requires that the marginal social *net benefit* of income be the same for every household. This net benefit is the marginal welfare of the marginal utility of income to the household *less* the marginal welfare cost of the additional consumption that the household would carry out as the result of an addition to income.

Wildasin states the following propositions for the many-person case:

1. Simple neutrality implies the Marginal Revenue Proportionality Rule.
2. Simple neutrality need not imply the Ramsey Rule.
3. Extended neutrality need not imply the Marginal Revenue Proportionality Rule.
4. Extended neutrality implies the Ramsey Rule.

In the next section I show that Proposition 3 is false.

II. Extended Neutrality and the Samuelsonian Consumer

Wildasin obtained his extended neutrality condition by supposing that, in addition to commodity taxation, the government can allocate lump sums s_0^h of the numeraire to households where

$$\sum_h s_0^h = 0$$

He derived the extended neutrality condition from the following equations:

$$(5) \qquad q\partial x^h/\partial s_0^h = 1 \qquad h=1,\ldots,H$$

$$(6) \qquad \partial L/\partial s_0^h = 0 \qquad h=1,\ldots,H$$

where L is the Lagrangian:

$$L = W(\{v^h\}) + \rho(y^f + y^g - X) - \lambda^f \phi^f y^f + \theta \sum_h s_0^h$$

Since (5) always holds if $qx^h = s_0^h$, the derivation can be reversed to show that the extended neutrality condition implies that (6) holds. Therefore, a solution to the optimal tax problem (1) which also satisfies extended neutrality is also a solution to the problem:

$$(7) \quad \max_{q,y^f} \left\{ \max_{\{s_0^h\}} \left[W(\{v^h(q,s_0^h)\}) : \sum_h s_0^h = 0 \right] \right. \left. y^f + y^g = \sum_h x^h, \quad \phi^f y^f \leqslant 0 \right\}$$

Paul Samuelson has shown that if, at every price, income is redistributed so as to maximize social welfare, then the resulting aggregate demand functions can be obtained from budget maximization with respect to a "utility function" over aggregate consumption. Therefore problem (7) is equivalent to the optimal tax problem with a single consumer, and all tax rules which are valid for an economy with a single consumer are also valid when interpreted in terms of the demands of the Samuelsonian "aggregate consumer." In particular, contrary to Wildasin's Proposition 3, the marginal revenue proportionality rule holds.[1]

In his discussion of the extended neutrality condition Wildasin (p. 894) explicitly rejects the notion that the government is carrying out lump sum transfers *in conjunction* with its choice of commodity taxation —as is implied in problem (7). His interpretation is rather that the welfare maximizer would *choose* not to carry out any lump sum redistribution of income, even if empowered to so do. However, these differences in interpretation have no bearing on the mathematical fact that, as stated above, a solution to the optimal tax which satisfies extended neutrality is also a solution to problem (7). This mathematical fact is all that is required to show that the marginal revenue proportionality rule holds.

The suspicion might linger that the redistribution of lump sums implicit in problem (7) might in some way interact with the

[1] The Gorman-Nataf conditions also imply the existence of a Samuelsonian consumer and hence this rule, as Wildasin notes.

choice of q,y^f so that Samuelson's result is not applicable. To allay this suspicion, we now show *directly* that problem (7) can be interpreted as an optimal tax problem with a single consumer. In so doing we effectively prove Samuelson's result without the smoothness and curvature assumptions on W and u^h which he used but which are in fact unnecessary.[2] All we assume is that $W(\{u^h(x^h)\})$ has a maximum over any compact set in $R^{H(n+1)}$.

Problem (7) is equivalent to each of the following problems:

$$\max_{q,y^f}\left\{\max_{\{s_0^h\}}\left[\max_{\{x^h\}}\left[W(\{u^h(x^h)\}):qx_h=s_0^h\right]\right.\right.$$

$$\left.\left.\sum_h s_0^h=0\right]y^f+y^g=\sum_h x^h,\phi^f y^f\leqslant 0\right\}$$

$$\max_{q,y^f}\left\{\max_{\{x^h\}}\left[W(\{u^h(x^h)\}):q\sum_h x^h=0\right]\right.$$

$$\left.y^f+y^g=\sum_h x^h,\phi^f y^f\leqslant 0\right\}$$

$$\max_{q,y^f}\left\{\max_X\left[\max_{\{x^h\}}\left[W(\{u^h(x^h)\}):\sum_h x^h\right.\right.\right.$$

$$\left.\left.\left.=X\right]qX=0\right]y^f+y^g=X,\phi^f y^f\leqslant 0\right\}$$

In the last problem, let $U(X)$ be the result of the maximization in the innermost square bracket. $U(X)$ is utility function of the "Samuelsonian consumer," defined over aggregate consumption. Then problem (7) is equivalent to the problem:

$$\max_{q,y^f}\left\{\max_X\left[U(X):qX=0\right]\right. \tag{8}$$

$$\left.y^f+y^g=X,\phi^f y^f\leqslant 0\right\}$$

This is just the optimal tax problem in an economy with a single consumer with utility function $U(X)$ defined over the consumption vectors X. This vindicates my earlier statement that the extended neutrality condition implies that all tax rules which are valid in an economy with a single consumer are also valid when interpreted in terms of the demands of the Samuelsonian consumer.[3]

Given the first-order conditions (2), the marginal revenue proportionality rule is equivalent to the following statement:

$$\sum_h W_h v_I^h x_k^h / X_k \text{ is independent of } k \tag{9}$$

We can give an intuitive explanation for (9) in terms of the indirect utility function of the Samuelsonian consumer:

$$V(q,I)=\max_X\left[U(X):qX=I\right]$$

Since problems (7) and (8) are equivalent, $V(q,0)$ equals the result of the maximization inside the outermost brackets in (7). Taking derivatives with respect to q_k we therefore have

$$\sum_h W_h v_k^h = V_k \qquad k=0,\ldots,n$$

Applying Roy's formula to the terms v_k^h and V_k we conclude that the expression in (9) equals V_I for all k.

Several authors[4] have noted that (9) holds if $W_h v_I^h \equiv \mu$. Wildasin's advance is to distinguish this condition (simple neutrality) from the marginal condition for an optimal distribution of income (extended neutrality) and to point out that the Ramsey Rule fails if only simple neutrality holds.[5] Another significant contribution is to prove that extended neutrality implies the Ramsey Rule, stated in terms of the compensated demands of *individual* agents: my analysis only yields

[2] The proof of Samuelson's result implicit in the following argument arose from discussions with John Fountain.

[3] A more direct, but less intuitive proof can be obtained as follows. Write down the Lagrange equations associated with each of the five maximization problems from problem (7) to problem (8). With elementary manipulations, (2), (5), and the extended neutrality condition can be shown to imply each of these sets of Lagrange equations in succession. From the last set, the tax rules corresponding to those for a one-consumer economy can be obtained.

[4] See Herbert Mohring, Martin Feldstein, and H. A. J. Green. Green in fact states incorrectly that $W_h v_I^h \equiv \mu$ holds if there is an optimal distribution of wealth (income in Wildasin's terms).

[5] Peter Diamond also emphasizes the difference between these two marginal conditions.

the Ramsey Rule in terms of the compensated demands of the Samuelsonian consumer. Wildasin failed to see, however, that extended neutrality also implies (9) and hence the Marginal Revenue Proportionality Rule.

III. Conclusions

If revenue must be raised by commodity taxation, and optimality is defined in terms of the welfare function appearing in the condition for extended neutrality, then this condition implies that all the tax rules obtained for a one-consumer economy are necessary conditions for optimal taxes in the corresponding many-consumer economy, when appropriately interpreted. To this extent, the extended neutrality assumption *is* an "...unambiguous way to segregate the distributional and efficiency aspects of the optimal tax problem" contrary to Wildasin's conclusion (p. 896).

REFERENCES

P. Diamond, "A Many-Person Ramsey Tax Rule," *J. Publ. Econ.*, Nov. 1975, *4*, 335–42.

M. S. Feldstein, "Distributional Equity and the Optimal Structure of Public Prices," *Amer. Econ. Rev.*, Mar. 1972, *62*, 32–36; "Errata," Sept. 1972, *62*, 763.

H. A. J. Green, "Two Models of Optimal Pricing and Taxation," *Oxford Econ. Pap.*, Nov. 1975, *27*, 352–82.

H. Mohring, "The Peak Load Problem with Increasing Returns and Pricing Constraints," *Amer. Econ. Rev.*, Sept. 1970, *60*, 693–705.

P. A. Samuelson, "Social Indifference Curves," *Quart. J. Econ.*, Feb. 1956, *70*, 1–22.

D. E. Wildasin, "Distributional Neutrality and Optimal Commodity Taxation," *Amer. Econ. Rev.*, Dec. 1977, *67*, 889–98.

Distributional Neutrality and Optimal Commodity Taxation: Reply

By David E. Wildasin*

In a way Leslie Young's comment on my earlier paper is puzzling. My Proposition 3 is based on equation (19) which, reproduced here as (1′), shows that optimal taxation requires

$$(1') \quad (1-\theta)-\sum_h \frac{x_k^h}{X_k}\frac{\partial(tx^h)}{\partial I^h}=\frac{\partial(tX)/\partial t_k}{X_k}$$

given the assumption of extended neutrality. If (1′) is valid, it is obvious that the Marginal Revenue Proportionality Rule[1] (*MRPR*) *cannot* obtain unless $\Sigma_h(x_k^h/X_k)(\partial(tx^h)/\partial I^h)$ is independent of k, and it is just as obvious that this condition holds only under special assumptions on consumer preferences. The proof of Proposition 3 is ironclad given (1′), so the obvious (and only) way to disprove it is to expose an error in the derivation of (1′). Young neither proves nor asserts that (1′) is false; in fact it is not.

On the other hand, Paul Samuelson did show that optimal redistribution does lead to aggregate demand behavior that is consistent with the restrictions that demand theory places on a single utility-maximizing consumer, and Young is indeed correct in his belief that this allows one to establish the single person results (both *MRPR* and the Ramsey Rule (*RR*) characterize the optimal tax structure) in the many-person case. This does not contradict my earlier results, however, and Young's contention to the contrary rests on a misunderstanding of the nature of the optimal tax problem he discusses. To be sure that neither my results nor those that Young purports to establish are misinterpreted, I propose to show in some detail how the two optimal tax rules can be derived for the many-person economy on the assumption that optimal lump sum redistribution of a certain kind is *actually performed.* I will also show that the assumption that such redistribution takes place—which is crucial to the conclusion that Young seeks to establish—alters in a fundamental way the problem that I originally investigated. Thus Young's conclusion differs from mine not because my results are incorrect, but because he considers a different problem.

To get on with the analysis, which presumes familiarity with the earlier discussion, let there first of all be given an *arbitrary* social welfare function (*SWF*) W, defined over individual utility levels. Recall that households are assumed to maximize utility by choice of x^h subject to

$$(1) \qquad qx^h=s_0^h$$

where q and s_0^h are taken as parametrically given. This yields the demand functions $x^h(q,s_0^h)$ and also the indirect utility functions $v^h(q,s_0^h)$.

Now suppose that there is an agent who, in Young's words, makes sure that "at every price, income is redistributed so as to maximize social welfare" (p. 234). This agent then takes consumer prices q as given and chooses a vector $s_0=(s_0^1,\dots,s_0^H)$ to

$$(\mathrm{R}) \quad \max_{s_0} W\left(v^1[q,s_0^1],\dots,v^H[q,s_0^H]\right)$$

subject to

$$(2) \qquad \sum_h s_0^h=S$$

*Department of economics, Indiana University. This material is based on work supported by the National Science Foundation under grant SOC78-05195.

[1]Interpreted in the natural way—that is, in terms of the sum of the derivatives of the ordinary individual demand functions (which is what we would actually observe in the marketplace in the absence of Youngian redistribution), as discussed in detail below.

where S is some fixed parameter. Ordinarily $S \equiv 0$ so (R) is a pure redistribution problem. The first-order conditions associated with this problem are (2) and

$$(3) \qquad W_h v_I^h - \mu = 0 \qquad h = 1, \ldots, H$$

where μ is the multiplier associated with (2). Assuming that $W(\{v^h(q, s_0^h\})$ is strictly quasi concave in s_0 for any given q,[2] one can solve for (μ, s_0) from (2) and (3) as continuously differentiable functions of (q, S). Then define

$$(4) \qquad V(q,S) = W(\{v^h(q, s_0^h[q,S])\})$$

and

$$(5) \qquad X(q,S) = \sum_h x^h(q, s_0^h[q,S])$$

Then the problem

$$(\mathrm{P}) \qquad \max_{q, y^f} V(q,0)$$

subject to

$$(6) \qquad y^f + y^g - X(q,0) = 0$$

and

$$(7) \qquad \phi^f y^f \leqslant 0$$

is equivalent to Young's (7), and is the natural way to exploit the Samuelsonian representative consumer idea in trying to recast the many-person tax problem in a single person mold. The functions $V(q,S)$ and $X(q,S)$ are of course to play the role of indirect utility and demand functions for this representative consumer.

Using definition (4) and differentiating (2), we have

$$(8) \qquad V_S = \sum_h W_h v_I^h \frac{\partial s_0^h}{\partial S} = \mu$$

[2]This is not a terribly strong requirement. Young believes that one can dispose with it, but it seems clear that it is necessary if one wishes to derive optimal tax rules, such as (16) below, involving the derivatives of aggregate demand functions.

$$(9) \qquad V_i \underset{\mathrm{def}}{=} \frac{\partial V}{\partial q_i} = \sum_h W_h v_i^h + \sum_h W_h v_I^h \frac{\partial s_0^h}{\partial q_i}$$
$$= \sum_h W_h v_i^h$$

It follows that

$$(10) \qquad \frac{V_i}{V_S} = -X_i$$

From (5) we have

$$(11) \qquad \frac{\partial X}{\partial S} = \sum_h \frac{\partial x^h}{\partial I^h} \frac{\partial s_0^h}{\partial S}$$

$$(12) \qquad \frac{\partial X}{\partial q_i} = \sum_h \frac{\partial x^h}{\partial q_i} + \sum_h \frac{\partial x^h}{\partial I^h} \frac{\partial s_0^h}{\partial q_i}$$

Also, using (1) and (5), one has

$$(13) \qquad X_i + q_i \frac{\partial X}{\partial q_i} = 0$$

$$(14) \qquad q \frac{\partial X}{\partial S} = 1$$

It is now straightforward that a solution to (P) is characterized by

$$(15) \qquad \frac{\mu}{\rho_0} X_k = t \frac{\partial X}{\partial q_k} + X_k \qquad \text{all } k$$

where ρ_0 is the Lagrange multiplier associated with the 0th of equations (6). The marginal revenue proportionality result

$$(16) \qquad \frac{\mu}{\rho_0} = \frac{\partial(tX)/\partial t_k}{X_k} \qquad \text{all } k$$

follows immediately. Note that the derivatives of X appearing in (16) are *not* simply the pure price derivatives $\Sigma_h \partial x^h / \partial q_k$, as can be seen by recalling (12).

Now rewrite (16) as

$$(17) \qquad \frac{\mu}{\rho_0} = 1 + \frac{t}{X_k}\left(\frac{\partial X}{\partial t_k} + X_k \frac{\partial X}{\partial S}\right) - t \frac{\partial X}{\partial S}$$
$$= 1 + \sum_i \frac{t_i S_{ik}}{X_k} - t \frac{\partial X}{\partial S}$$

where we define $S_{ik} = \partial X_i/\partial t_k + X_k(\partial X_i/\partial S)$. On the assumption that $S_{ik} = S_{ki}$, simple rearrangement yields

$$(18) \qquad \frac{\mu}{\rho_0} - 1 + t\frac{\partial X}{\partial S} = \sum_i \frac{t_i S_{ki}}{X_k}$$

which is precisely the Ramsey Rule, with the S_{ik}'s playing the role of substitution terms.

I shall now work out an implicit solution for s_0 as a function of q and S for a simple special case and then show directly that $S_{ik} = S_{ki}$. Suppose $H=2$ and suppose the social welfare function is

$$(19) \qquad W = \alpha_1 v^1 + \alpha_2 v^2$$

If v^1 and v^2 are chosen so that the marginal utility of income diminishes (i.e., $v^h_{II} < 0$), then we can use (2) and (3) to solve for $s_0(q,S)$. This condition is met, for example, if the direct utility functions are of the familiar form

$$u^h = \beta^h_0 \log(x^h_0 + \omega^h) + \sum_{i=1}^{n} \beta^h_i \log x^h_i$$

with

$$\sum_{i=0}^{n} \beta^h_i = 1, \omega^h > 0 < \beta^h_i \qquad \text{all } i$$

It is straightforward to find

$$(20) \qquad \frac{\partial s^1_0}{\partial S} = \frac{\alpha_2 v^2_{II}}{\alpha_1 v^1_{II} + \alpha_2 v^2_{II}}, \qquad \frac{\partial s^2_0}{\partial S} = 1 - \frac{\partial s^1_0}{\partial S}$$

$$(21) \qquad \frac{\partial s^1_0}{\partial q_i} = \frac{\alpha_2 v^2_{Ii} - \alpha_1 v^1_{Ii}}{\alpha_1 v^1_{II} + \alpha_2 v^2_{II}}, \qquad \frac{\partial s^2_0}{\partial q_i} = -\frac{\partial s^1_0}{\partial q_i}$$

$$i = 1, \ldots, n$$

where $v^h_{Ii} = \partial v^h_I/\partial q_i$. Using $v^h_{Ii} = -v^h_I(\partial x^h_i/\partial I^h) - x^h_i v^h_{II}$, (3), and (20), we can write

(22)

$$\frac{\partial s^1_0}{\partial q_i} = \frac{\mu(\partial x^1_i/\partial I^1 - \partial x^2_i/\partial I^2)}{\alpha_1 v^1_{II} + \alpha_2 v^2_{II}} + x^1_i - X_i\frac{\partial s^1_0}{\partial S}$$
$$= \gamma\left(\frac{\partial x^1_i}{\partial I^1} - \frac{\partial x^2_i}{\partial I^2}\right) + x^1_i - X_i\frac{\partial s^1_0}{\partial S}$$

Letting $D_i = \partial x^1_i/\partial I^1 - \partial x^2_i/\partial I^2$ for brevity, using (20) and (21) to eliminate derivatives of s^2_0, and using (11) and (12), we can write out S_{ik} in detail:

$$(23) \qquad S_{ik} = \left.\frac{\partial x^1_i}{\partial q_k}\right|_{\bar{u}} + \left.\frac{\partial x^2_i}{\partial q_k}\right|_{\bar{u}} - x^1_k\frac{\partial x^1_i}{\partial I^1}$$
$$- x^2_k\frac{\partial x^2_i}{\partial I^2} + D_i\frac{\partial s^1_0}{\partial q_k} + X_k\left[D_i\frac{\partial s^1_0}{\partial S} + \frac{\partial x^2_i}{\partial I^2}\right]$$

Now using (22) it follows that

$$(24) \qquad S_{ik} = \left.\frac{\partial x^1_i}{\partial q_k}\right|_{\bar{u}} + \left.\frac{\partial x^2_i}{\partial q_k}\right|_{\bar{u}} + \gamma D_i D_k = S_{ki}$$

using the symmetry of the individual substitution terms. We have established the symmetry of the substitution terms of the price derivatives of the representative consumer's demand function; the Ramsey Rule therefore characterizes an optimal tax structure, as shown in (18). Note from (24), however, that the S_{ik}'s are *not* the sum of the substitution terms of the individual consumers. We can summarize all this in the form of a

THEOREM: *Given lump sum redistribution of the type described in problem* (R), *an optimal tax structure—that is, a tax structure emerging from the solution of problem* (P)*—can be characterized by both MRPR and RR, these tax rules being interpreted in terms of the demand functions of the representative consumer.*

Does this valid theorem contradict any of my earlier results? No, because I did not allow for—that is, I implicitly ruled out —the Samuelsonian redistribution, embodied in the functions $s^h_0(q,S)$ obtained as solutions to problem (R), that is so crucial to the proof of the above theorem. To see how crucial this redistribution really is, recall that the "ordinary" and "compensated" price derivatives of the representative demand function that appear in (16) and (18) are *not* the sum of the corresponding derivatives of the individual demand functions

$x^h(q,0)$ of consumers who have lump sum incomes fixed identically at zero, independent of q—yet the latter are what would be observed in the absence of redistribution.

Young believes that the problem (P) is equivalent to the standard problem

$$\text{(P}'\text{)} \qquad \max_{q,y^f} W(\{v^h(q,0)\})$$

subject to (7) and

$$\text{(25)} \qquad y^f + y^g - \sum_h x^h(q,0) = 0$$

when the SWF exhibits extended neutrality. He argues that extended neutrality implies that $s_0^h = 0$, all h, solves (R), and that the solution set to (P′), in which each s_0^h is in effect constrained to be identically zero, will then coincide with the solution set to (P). To clear up a minor difficulty at first, note that simple—not extended—neutrality is needed to have $s_0^h = 0$ at a solution to (R). This is obvious from the first-order conditions (3), and should immediately alert the reader to the fact that Young has not shown what he thinks he has. But suppose that W does satisfy the simple neutrality condition at a solution $(\tilde{q}, \tilde{y}^f)$ to problem (P), so that the functions $s_0^h(q,0)$ giving the solution to (R) have the property that $s_0^h(\tilde{q},0) = 0$ all h. Is $(\tilde{q}, \tilde{y}^f)$ then a solution to (P′)? No, because the proper choice of consumer prices in the problem (P) must take into account, *at the margin*, the effects of prices on lump sum incomes, whereas this is not true in the case of the problem (P′). The fact that, under simple neutrality, the functions $s_0^h(q,0)$ have a value of zero at a solution to (P) does *not* imply that the *derivatives* of these functions also vanish, and (20) and (22) show that the derivatives will not in fact vanish, in general.[3] But as shown by (16) and (18), interpreted in view of (12) and (23), these derivatives play a critical role in the determination of a solution to (P), and in the derivation of RR and $MRPR$. In the case of problem (P′), when redistribution is ruled out ($s_0^h \equiv 0$ all h), the aggregate demand functions are just $\Sigma_h x^h(q,0)$, which, unlike the $X(q,0)$ functions, do not have the properties of individual demand functions; specifically

$$\partial\Big(\sum_h x^h[q,0]\Big)/\partial q_k = \sum_h (\partial x^h/\partial q_k)_{\bar{u}} - \sum_h x_k^h(\partial x^h/\partial I^h)$$

$$\neq \sum_h (\partial x^h/\partial q_k)_{\bar{u}} - \Big(\sum_h x_k^h\Big)\sum_h (\partial x^h/\partial I^h)$$

that would be required for both RR and $MRPR$ to hold. This, of course, is just what I originally showed in my paper.[4]

[3]As Samuelson observed, "a stipulated percentage breakdown of income [for example, Papa 10 percent, Mama 51 percent, Junior 39 percent] cannot be an optimal rule for a nonshibboleth social welfare function." Rather, "[i]ncome must always be reallocated among the members of our family society so as to keep the 'marginal social significance of every dollar' equal" (p. 11), in the face of changing prices. In other words, $\partial s_0^h/\partial q_i \neq 0$.

[4]It might be helpful to clarify the relationship between the kind of redistribution that I discussed on pp. 893–94 of my original paper in motivating the extended neutrality definition, and the kind of redistribution implied by Young's approach and described above in problem (R). Young obviously believes the two are equivalent. That they are actually different is most easily seen by comparing the conditions for optimal redistribution in each case: equation (18) in my original paper for the redistribution considered there, (3) above for the redistribution considered by Young. In contrast to the latter, the redistribution that I discussed takes into account the effect of redistribution on government revenue via the term $\partial(tx^h)/\partial I^h$ appearing in my original (18). (As a footnote to the footnote, Young errs in interpreting $\partial(tx^h)/\partial I^h$ as "the marginal welfare cost of the additional consumption that the household would carry out as the result of an addition to income" (p. 234) which must be *subtracted* from $(W_h v_I^h)\rho_0$ to arrive at marginal social *net* benefit of income. To the extent that additional income results in additional demand for tax-distorted goods, a marginal welfare *gain* is generated that must be *added* to $(W_h v_I^h)/\rho_0$.) When accompanied by optimal selection of (q,y^f), redistribution of the type that I originally discussed thus leads to a welfare level higher than that attained by optimal choice of (q,y^f) with *endogenous* consumer price-contingent redistribution as set out in problem (P). Thus the redistribution of problem (R) is optimal in a somewhat restricted sense: rather than requiring the redistribution program to satisfy (3) at every price, it would be better to choose prices, output, and transfers *simultaneously*.

In sum, Young is correct in his view that Samuelsonian redistribution does lead consumers, in the aggregate, to behave like a single consumer, and that both *MRPR* and *RR* characterize an optimal tax structure in this case. He fails to appreciate the fact that this redistribution alters the optimal tax problem in a fundamental way, so that rather than contradicting my original results, the new problem leads to new results.

Since redistribution must actually be carried out for the theorem stated above to be valid, its scope of application is presumably nil. This does not mean that it is uninteresting, however, for the thought of using a Samuelsonian representative consumer to analyze the optimal tax problem in the many-person economy is a natural one, the implications of which are worth exploring. Young deserves credit for suggesting this approach. Nothing in Young's comment or this reply would, however, lead me to change the statement that "[t]here is simply no unambiguous way to segregate the efficiency and distributional aspects of the optimal tax problem" (p. 896).

REFERENCES

P. A. Samuelson, "Social Indifference Curves," *Quart. J. Econ.*, Feb. 1956, *70*, 1–22.

D. E. Wildasin, "Distributional Neutrality and Optimal Commodity Taxation," *Amer. Econ. Rev.*, Dec. 1977, *67*, 889–98.

L. Young, "Distributional Neutrality and Optimal Commodity Taxation: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 233–36.

Who Benefits from Economic Development?: Comment

By MONTEK S. AHLUWALIA, JOHN H. DULOY, GRAHAM PYATT, AND T. N. SRINIVASAN*

According to Gary Fields in a recent article in this *Review*, the average income of the poor in Brazil grew by some 63 percent in real terms between 1960 and 1970; in contrast, the average income of the nonpoor grew by about 28 percent. In recognizing that these results depend on some particular assumptions, Fields claims that "...under *no* possible alternative assumptions would ...[this] conclusion about the relative rates of income growth among the poor and nonpoor, be reversed qualitatively" (p. 573 fn. 6). This note takes issue with Fields' analysis on three counts. First, the relevance of the data underlying the analysis is tenuous for the issues being addressed. Second, the logic of Fields' analysis is wrong, and his numerical results are impossible. And third, correct logic leads to the conclusion that a wide range of numerical results is feasible so that strong conclusions cannot be drawn. Setting aside the limitations of the data, the only certainty is that either Fields understates the number of poor in 1970, or he overstates their income growth, or both. Subsequent sections treat each of these three issues in turn.[1]

I. Data Limitations

The data used by Fields were analyzed earlier in articles by Albert Fishlow and Carlos Langoni. They relate to the distribution of the economically active population across income classes in 1960 and 1970. These data (based on responses to a single question in a demographic census) have not been checked for consistency with national accounts. It is known that the exclusion of nonmonetary income and understatement of nonsalary incomes in the census results in an underestimate of the mean and also alters the distribution. The accuracy and comparability of these data are therefore open to doubt. These issues are familiar enough and will not be repeated here despite their importance. Rather, we draw attention to two points which go beyond these familiar problems.

The first concerns price deflators. Fields estimates a price index of 3,532 for 1970 with a base of 100 for 1960. This same value is used to deflate all income class boundaries in 1970 to be comparable with 1960 figures. At no point is it recognized that the appropriate price index may differ by income levels. Yet it is obviously the case that the consumption patterns of rich and poor in Brazil are different, implying that a single index of inflation is potentially inappropriate across the full range of the income scale.[2]

The other issue concerns the relevance of individual earnings data for analyzing poverty. While there clearly is a link between the two, poverty is primarily a matter of household living standards. The distribution of households according to their living standards would depend on the joint distribution of the following household characteristics in particular: age-sex distribution of its members and the position of its earning members in the income distribution among earners. Knowledge of this joint distribution and its changes over time is essential if we

*Development Research Center, World Bank. Views expressed are ours alone and do not necessarily reflect those of the World Bank. We thank Gary Fields for helpful discussions on an earlier draft.

[1]The logic of Fields' analysis has been queried previously in private correspondence by us, among others. In his original paper Fields cites a response to these queries.

[2]These comments on the data and on the uniform deflation across income classes do not apply only to the work of Fields; they also apply with equal force to the work of Fishlow and Langoni.

TABLE 1—LOWER END OF THE SIZE DISTRIBUTION OF INCOME FOR THE ECONOMICALLY ACTIVE POPULATION: BRAZIL 1960 AND 1970[a]

Monthly Income NCr$ 1960	Data Points 1960		Data Points 1970		Fields' Interpolation 1970		
	Share of: Population	Share of: Income	Share of: Population	Share of: Income	Share of: Population	Share of: Income	Implied Mean Income
None	14.7	0.0	11.7	0.0	11.7	0.0	0
0–2.1	22.3	5.2	31.7	8.0	23.8	6.0	1.84
2.1–2.8	–	–			7.9	2.0	1.84

Source: Fields, Table 1 and text.
[a]Shown in percent.

are to use the earnings distribution for an evaluation of the extent of poverty and its change over time. These matters are not addressed in Fields' paper.

II. The Logic of the Calculations

Having deflated the 1970 distribution of earnings to 1960 prices as discussed above, Fields' major numerical problem derives from the fact that the income class intervals for the 1960 and 1970 distributions of income are not the same. Table 1 above poses the problem. It shows that for both 1960 and 1970, frequencies at zero income are observable. However, the next (monthly) income bracket for 1960 is from zero to NCr$2.1, while for 1970 the bracket is from zero to NCr$2.8. Since Fields defines the poor as those with income below NCr$2.1, his formal problem is to estimate the population proportion below this level in 1970, and the proportion of income they receive. His estimates are shown in Table 1 as 35.5 percent for population (11.7+23.8) and 6.0 percent for income. Hence his estimate of a 63 percent growth in real income of the poor is obtained.[3]

[3]Average income of the poor in 1960 is NCr$0.77. For 1970, the estimate can be obtained by multiplying average income of the population (NCr$7.3) by the ratio of 6.0 to 35.5. This gives a figure of NCr$1.23 as the average income of the poor, i.e., an increase of 60 percent. The difference between this figure and the 63 percent quoted in Fields' text is presumably due to rounding errors: Fields quotes NCr$1.3 as average income of the poor in 1970 and NCr$0.8 as the comparable 1960 figure. The ratio 1.3 to 0.8 does in fact give an increase of 62.5 percent over the decade.

Fields' estimates of population and income proportions of the poor in 1970 are logically impossible. This can be seen from the last column of Table 1 where the mean incomes of each class interval (as implied by Fields' interpolations) are given. Obviously, the 7.9 percent of population having incomes between NCr$2.1 and 2.8 cannot have a *mean* income of NCr$1.84!

III. A Correct Formulation

The clue to understanding Fields' logical error and how to correct it lies in realizing that given the data points for 1970 in Table 1, the percentage P of the poor (i.e., those having incomes not exceeding NCr$2.1) and their income share π have to satisfy the constraints that (i) the mean income of the poor excluding those with zero incomes does not exceed NCr$2.1, and (ii) the mean income of the nonpoor in the class interval (0,2.8) (i.e., those with incomes exceeding NCr$2.1 but not exceeding NCr$2.8) lies between NCr$2.1 and NCr$2.8.

The first requirement leads to

$$\frac{7.3\pi}{(P-11.7)} \leqslant 2.1 \tag{1}$$

since 11.7 percent of the poor have no incomes and the mean income of the entire population is NCr$7.3. The second requirement leads to

$$2.1 < \frac{7.3(8-\pi)}{(43.4-P)} \leqslant 2.8 \tag{2}$$

since the income share of the class interval (0, 2.8) is 8 percent and the share of population having incomes not exceeding NCr$2.8 is 43.4 percent. It can be easily verified that (1) is implied in the left-hand side of (2).

Rewriting (2) we get

$$(2') \quad 0.38P - 8.65 \leqslant \pi < 0.29P - 4.48$$

Fields' point $P = 35.5$, $\pi = 6.0$ satisfies the left inequality, but not the right inequality. The left inequality is represented by the straight line BD and the right by BC in Figure 1. The feasible poverty point has to lie in the triangle CBD.[4] Clearly, Fields' point lies outside this feasible triangle. Thus, if we accept Fields' population share of 35.5 percent for the poor, their income share π is *overstated* by him. On the other hand, if we accept his income share of 6 percent for the poor, their population share is understated by him. It is clear from Figure 1 that the data allow a wide range of values of P and π.

We can push the analysis a bit further. The mean income of the poor relative to the mean income of the nonpoor in 1970 is by definition $\pi(100-P)/P(100-\pi)$. In 1960 this ratio was .0934. Thus for the mean income of the poor to have grown faster than that of the nonpoor between 1960 and 1970 we must have

$$(3) \quad \frac{\pi(100-P)}{P(100-\pi)} > .0934$$

or

$$\pi > \frac{9.34P}{(100-P)+.0934P}$$

The curve obtained by replacing the inequality in (3) by equality is $OXYZ$ in Figure 1. Clearly any feasible poverty point that lies above this curve will be consistent with the qualitative conclusion of Fields, namely, the mean income of the poor grew faster than that of the nonpoor. By the same token, any feasible poverty point below the curve will lead to the opposite conclusion. Without additional information, we cannot infer whether the "actual" poverty point fell in the former or latter region. If we set P equal to Fields' estimate of 35.5, then the lower bound on π is 4.84 which places (P, π) *below* the curve $OXYZ$ whereas the upper bound on π (5.82) will place it above the curve. However, if we confine our attention to the bottom 37 percent of the economically active population (the same proportion that constituted the poor in 1960) and arbitrarily treat them as the poor in 1970 (i.e., if we set $P = 37$), *then* all feasible income shares π associated with this P imply that the mean income of the poor grew faster than that of the nonpoor. But this is an entirely arbitrary assumption which implies that the proportion of the population who are poor was neither decreased nor aug-

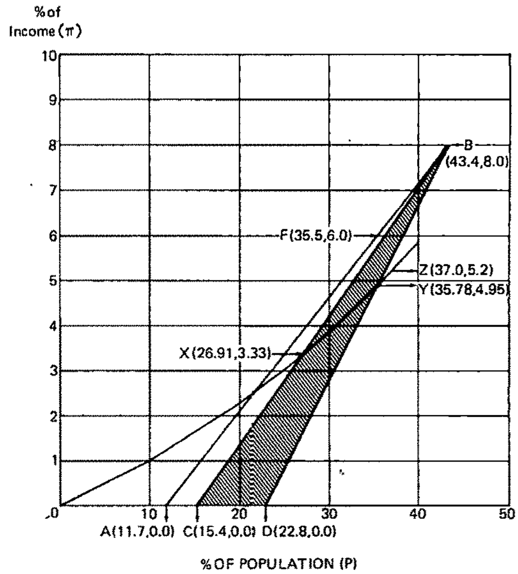


FIGURE 1

[4]The upper bound of the feasible space of the Lorenz curve is defined by its chord AB. The lower bound is given by the two tangents at point A (line OD) and at point B (line BD). The feasible space for the poverty point on the Lorenz curve depends upon the value of poverty income assumed. For Fields' poverty line of NCr$2.1, this additional constraint is represented by the line BC.

mented by the process of aggregate income growth over the decade 1960 to 1970.

To summarize, the data do not sustain the strong conclusions contained in Fields' paper on the growth of income of the poor as defined by him. More importantly, we would emphasize that the data themselves are not suitable for analysis of changes in poverty. For this reason we have refrained from any detailed discussion of the substantive issues, both conceptual and factual, concerning the impact of Brazilian economic growth on the poor.

REFERENCES

G. S. Fields, "Who Benefits from Economic Development?—A Reexamination of Brazilian Growth in the 1960's," *Amer. Econ. Rev.*, Sept. 1977, *67*, 570–82.

A. Fishlow, "Brazilian Size Distribution of Income," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 391–402.

C. Langoni, "Distribuiçao da Renda e Desenvolvimento Econômico do Brasil," *Estudos Econômicos*, Oct. 1972, *2*, 5–88.

Who Benefits from Economic Development?: Comment

By PAUL BECKERMAN AND DONALD COES*

Gary Fields' recent examination in this *Review* of the distributive effects of Brazilian growth in the 1960's is a novel attempt to answer the controversial questions of who benefits from economic growth and by how much. In contrast to more conventional studies of relative inequality, which use such measures as the Gini coefficient, Fields examines changes in the absolute sizes of income shares received by different income classes in Brazil between 1960 and 1970. His principal qualitative conclusion is that despite an increase in relative inequality, "the poor *did* participate in the rapid economic growth of the decade" (p. 570).

To our knowledge, no serious student of recent Brazilian economic history has maintained that the growth of the 1960's did not provide at least some benefit to each major income class, including the poor. Indeed, in view of the magnitude of Brazil's growth, any other outcome would have been highly improbable. Given the extremely low level of income earned by Brazil's poorest in 1960, if no more than a small fraction of the aggregate increase in personal income had gone to the poor, their per capita income would have increased dramatically. Using Fields' data and definition of poverty, for example, the allocation of less than a sixth of Brazil's total per capita income increase over the decade to the third of the population defined as poor would have doubled their incomes.[1] Fields' calculation that the mean income of the Brazilian poor increased by 63 percent may therefore be interpreted as an indication of how limited their participation in the decade's growth actually was.

The novelty of Fields' absolute-shares analysis does not lie in its principal qualitative conclusion, which is likely to hold for any rapidly growing market economy; rather the major innovation that Fields' approach offers is its method of measuring the distributive impacts of income growth. Our purpose in this comment is to demonstrate that quantitative conclusions based on the absolute shares method are not robust to small variations in assumptions about the extent of real income increase accruing to the population as a whole. The fundamental problem is that the summary measures calculated using the absolute shares approach are highly sensitive to minor changes in measured real income. As we show below, this is not simply a theoretical objection. Slightly higher estimates of inflation in Brazil during the 1960's than that used by Fields substantially alter many of his conclusions. Using the change in the São Paulo cost-of-living index (*CLI*) between 1960 and 1970 rather than Fields' inflation measure, but a procedure in all other respects identical to his, we find that his method actually shows an increase in the proportion of the population that he classifies as poor.

The data for Fields' study are those of Albert Fishlow, who showed that relative inequality as defined by the Gini coefficient increased between 1960 and 1970, based on

*University of Illinois. (Beckerman is visiting assistant professor, Center for Latin American Development Studies, Boston University. Coes is visiting economist, Fundação Centro de Estudos do Comércio Exterior, Rio de Janeiro.) We thank Werner Baer, Paulo de Tarso Andre, and Robert Ferber for helpful advice. Any remaining errors of fact or judgement are of course our own.

[1] Fields concludes that the per capita income of the population as a whole increased from 5.5 to 7.3 real 1960 NCr$ between 1960 and 1970. Mean per capita income of the poor, who were 37 percent of the population in 1960, rose from .8 to 1.3 NCr$, and nonpoor mean per capita income from 8.3 to 10.3 NCr$. An increase of .8 NCr$, doubling mean poor income, would therefore have required $(.37\times.8)/1.8=.165$ percent of the per capita income increase.

TABLE 1

Monthly Income in 1960 NCr$ (a)	Percentage of Population 1960 (Fishlow) (b)	1970 (*CLI*) (c)	1970 (Fields) (d)	Cumulative Percentage 1960 (Fishlow) (e)	1970 (*CLI*) (f)	1970 (Fields) (g)
None	14.7	11.7	11.7	14.7	11.7	11.7
0.0–2.1	22.3	25.6	23.8	37.0	37.3	35.5
2.1–3.3	14.4	13.0	12.2	51.4	50.3	47.7
3.3–4.5	10.5	13.1	11.0	61.9	63.4	58.6
4.5–6.0	13.1	11.2	14.5	75.0	74.6	73.1
6.0–10.0	13.8	9.4	9.4	88.8	84.0	82.5
10.0–20.0	8.2	10.0	10.9	97.0	94.0	93.4
20.0–50.0	2.6	4.8	5.0	99.6	98.8	98.4
Over 50.0	0.5	1.2	1.6	100.1	100.0	100.0

data from the Brazilian censuses for those two years. As the two income distributions tabulated by Fishlow report frequencies by income bracket defined in current new cruzeiros (NCr$), the absolute shares method requires that the 1970 interval limits be deflated by an appropriate price index. Fields derives a deflator by dividing the ratio of nominal cruzeiro mean income in 1970 to that of 1960 by the corresponding ratio of mean incomes in 1970 and 1960 expressed in constant 1960 *U.S.* dollars, as reported by Fishlow, yielding a deflator of 35.32.

A number of deflators can be computed, however, from published Brazilian price indices. Although they vary little in their measurement of Brazil's inflation between 1960 and 1970, Fields' results are sensitive to the index chosen. In principle, the correct deflator for a study of change in personal incomes would be based on an appropriate national *CLI*. In the absence of such a series, we have used the São Paulo *CLI*.[2] On a 1965–67 base of 100, this index increased from a 6.22 monthly average in 1960 to 238 in 1970, yielding a deflator of 38.26 for the decade. In annual terms there is little difference between the two deflators: Fields' deflator is equivalent to an annual rate of 42.8 percent, while the deflator based on the São Paulo *CLI* is equivalent to an annual rate of 44.0 percent.

Our results using the São Paulo *CLI* are presented in columns (c) and (f) of Table 1. Save for the deflator used, we have followed Fields' approach in all respects, including the linear interpolation procedure he used to assign the 1970 brackets to brackets matching those of 1960. For the purpose of comparison, the distributions reported by Fields are shown in columns (d) and (g).

The most important feature of the 1970 distributions using the *CLI* as the deflator is that the cumulative percentage (col. (f)) is not lower in 1970 than in 1960 for all income brackets, as it is in Fields' study. The proportion of Brazil's economically active population earning less than 2.1 real 1960 NCr$ actually increased slightly during the decade, with an even larger increase in the proportion of the population earning less than 4.5 real 1960 NCr$.[3]

[2]The São Paulo *CLI* is compiled by the Bolsa de Mercadorias and the Instituto de Pesquisas Economicas of the Universidade de São Paulo. A historical series appears in *Conjuntura Econômica*, April 1977, Vol. 31. Fields' deflator is derived from Fishlow's study, which used the change in the implicit *GDP* deflator between 1960 and 1970.

[3]Much of the recorded gain in income by Brazil's poor between 1960 and 1970 may be due to increasing payment of farm workers in cash rather than in kind, due to the *Estatuto do Trabalhador Rural* (*Rural Labor Statute*) promulgated in the 1960's. This raises the possibility that the gain was more apparent than real. In response to this criticism by a referee, Fields reports in a footnote that the exclusion of the zero-income category and reestimation of the distribution does not alter his conclusions. Using the São Paulo *CLI*, we find that the resulting cumulative distribution shows an increase in the proportion of population below the poverty line, wherever it is drawn among the lower five income brackets. We note, moreover, that even the exclusion of the zero-income category does not resolve this problem, since many persons who reported some

TABLE 2

		CLI	Fields
A. Mean incomes, rates of growth, and relative measures of the "Poverty Gap"			
$\bar{Y}_p^{70}$	Mean poor income in 1970	1.1 NCr$	1.3 NCr$
$\bar{Y}_n^{70}$	Mean nonpoor income in 1970	10.1 NCr$	10.6 NCr$
$(\bar{Y}_p^{70}-\bar{Y}_p^{60})/\bar{Y}_p^{60}$	Rate of growth of poor mean income, 1960–70	38%	63%
$(\bar{Y}_n^{70}-\bar{Y}_n^{60})/\bar{Y}_n^{60}$	Rate of growth of nonpoor mean income, 1960–70	22%	22%
$\bar{Y}_n^{70}/\bar{Y}_p^{70}$	Nonpoor to poor mean income ratio in 1970	9.2	8.2
Poverty Gap in 1970 = (2.1 NCr$ − $\bar{Y}_p^{70}$) x	percentage of population below poverty line	37.3%	28.4%
Poverty Gap reduction in decade		19.9%	10.8%
B. Decomposition of total income growth 1960–70			
α	Enlargement effect = change in number of persons in high-income sector times income differential between sectors in base year	−2%	6%
β	Enrichment effect = change in income within high-income sector times number of people in sector in base years	91	82
γ	Interaction term	0	2
δ	Enrichment of low-income sector = change in income within the low-income sector times people in sector in base year	11	10
		100%	100%

In the second part of his paper, Fields presents a number of summary measures based on the absolute shares approach which are intended to compare the income change among the poorest groups with that of all others. Following Fishlow, who concluded (on the basis of minimum wage standards for the Northeast, Brazil's poorest region) that about 31 percent of the population in 1960 was below the minimum, Fields defines the "poverty line" as 2.1 1960 NCr$. As may be seen in Table 1, 37 percent of the economically active population fell below this line in 1960.

In Table 2 we report our calculations of various measures of the change in income of the poor and nonpoor during the 1960's, based on the São Paulo *CLI*, together with Fields' original results. Interested readers are referred to his article for the methods of calculation and additional data.

money income in 1960 and reported increases in money income in 1970 may have lost income in kind over the period. Monetization of rural income was encouraged by the *Rural Labor Statute*. On this point see Aluysio Sampaio.

As is clear from Table 2, the use of the São Paulo cost-of-living deflator rather than Fields' deflator has a substantial impact on all of the summary measures. In his discussion of his results, Fields suggests that the decomposition parameters α, β, and δ be computed for a number of countries and compared. The sensitivity of these parameters to small changes in measured real income growth suggests that such a comparison would be meaningless. Fields also notes that Brazil and the United States both reduced their respective poverty gaps by 41 percent in the decade.[4] Even accepting the arbitrary definitions of poverty in both nations implicit in the measures, it is obvious from Table 2 that this correspondence is really accidental.

[4]Fields calculates the 1960 poverty gap to be 48.1 percent, so that its 1970 level of 28.4 percent represents a 41 percent reduction.

Both Tables 1 and 2 demonstrate the sensitivity of income changes among different groups estimated by Fields' absolute shares method to minor variations in the assumed rate of real income growth. We do not argue that our estimates are any more reliable or informative than those of Fields; in fact, other price indices available would alter his estimates in an opposite direction.[5] One might argue, moreover, that two data points separated by a decade of extensive and highly variable economic change are hardly sufficient to bear the weight of the Brazilian income distribution controversy. Our point here is that the method itself cannot provide meaningful quantitative statements about changes in income earned by different groups, due to its sensitivity to small changes in the measured rate of real income growth.

In conclusion, we note that some readers might object to the assumption implicit in the absolute shares approach that the utility of each income earner is independent of every other recipient's income. The more traditional focus on relative inequality reflects the theoretically and empirically plausible assumption that the marginal utility of income is decreasing, while the absolute shares approach deliberately ignores this motive for the conventional preoccupation with worsening relative inequality. Even accepting the assumptions implicit in the absolute shares method, however, our results suggest that it does not produce quantitatively robust estimates of the distribution of the effects of economic growth.

[5]Other price indices for the period 1960 to 1970 reported in *Conjuntura Econômica* vary from an annual rate of 40.9 to 45.2 percent. Some critics of Brazilian economic policies would contend that official price indices have tended to overestimate the success of the government's struggle against inflation. In 1975, in fact, the Brazilian government revised upwards its own estimates of the rates of inflation for the years 1970–73.

REFERENCES

G. S. Fields, "Who Benefits from Economic Development?—A Reexamination of Brazilian Growth in the 1960's," *Amer. Econ. Rev.*, Sept. 1977, *67*, 570–82.

A. Fishlow, "Brazilian Size Distribution of Income," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 391–402.

Aluysio Sampaio, *Estatuto do Trabalhador Rural: Comentado*, São Paulo 1972.

Fundação Getúlio Vargas, *Conjuntura Econômica*, Rio de Janeiro, various issues.

Who Benefits from Economic Development?: Comment

By ALBERT FISHLOW*

Gary Fields' recent reexamination in this *Review* of the income distribution implications of Brazilian growth in the 1960's has attracted wide notice both in academic and official circles. I believe its apparent finding of substantial improvement for the poor is unfounded. My contribution is directed to clarifying three central points: 1) the inadequate nature of the data underlying Fields' inferences; 2) the extent to which relative deterioration of the income distribution in the 1960's actually did limit the absolute gains of the poor; and 3) the special character of Fields' emphasis on absolute incomes of the "poor" and "nonpoor" as a measure of welfare of the poor.

I

Fields' analysis starts from the uncorrected distributions of monetary income for the economically active populations in 1960 and 1970 that I initially derived in 1972, Tables 1 and 5. He then applies to that 1960 distribution a poverty line defined as a monthly income of NCr$2.1; that standard is defended by appeal to my finding that 31 percent of Brazilian families in 1960 had inadequate incomes:

> We must begin by establishing a poverty line. Something like 31 percent of Brazilian families were poor in 1960 by Brazilian definitions. Since it is not possible to identify those families exactly, we may suppose that those persons in the two lowest income brackets (i.e., less than 2.1 NCr$ constant), which in 1960 comprised 37.0 percent of the population, were below the poverty level. From now on, we will refer to these persons with incomes below 2.1 as "the poor" and the rest of the population as "the non-poor." [p. 573]

*Yale University.

Such a procedure is both critical to the findings and quite illegitimate. It ignores the difference between the distributions among individuals and families; it obscures the fact that the original analysis corrected for income in kind and regional price differences; and it takes too lightly the problems posed by inclusion of a large group of individual zero-income recipients.

Meaningful definition of poverty requires consumption, not recipient, units. After all, the most potent institution for income redistribution from the minority of income earners found in almost all societies is the family itself. Such consumption units, particularly in the agricultural sector, supplement their incomes by direct production for subsistence, employing young family members in the process. That is the clear explanation for the large numbers of economically active in Brazil reported without income: in 1960, 95 percent were in the agricultural sector; 71 percent were 19 or younger; 85 percent are in households in which the family head is 40 and above. The same sectoral concentration holds true for 1970. The census for that year additionally informs us that only 8 percent of the zero-income economically active were either unemployed or seeking work for the first time (see *Censo Demografico*, pp. 74–75). There is no basis for identifying such a zero-income group as the unemployed.

This group of family workers dominate Fields' "poor" in both 1960 and 1970. In each year they make up more than a third of the total. Moreover, the change in the number of poor between 1960 and 1970 is more than explained by the reduction in the number of unpaid family workers. That decline is of interest in itself, reflecting as it does increased commercialization in agriculture, a decline in the relative size of the agricultural population, and perhaps greater attraction to urban opportunities. That does

not tell us anything, however, about changes in the number or composition or average income of poor families over the decade, or about changes in unemployment of individual income recipients. Inclusion of the zero-income group as a proxy for other information is unjustified.

Simple exclusion is not the right procedure either. Exclusion, however, does make a dramatic difference to Fields' conclusions. Eliminating zero-income workers from the distribution fully reverses his quantitative and qualitative results. Fields' claim to the contrary (p. 573, fn. 7) is wrong, the result of arithmetic miscalculation. When zero-income workers are eliminated both in 1960 and 1970, the share of income received by the poor—defined by Fields as 35.9 percent of the economically active population receiving income—declines from 8.9 to 8.0 percent. That translates into an average rate of growth of absolute income for those under the poverty line of 19 percent, compared to 32 percent overall. The relative growth of income of the poor is quite likely overestimated, moreover, because of the procedures I had used to estimate the average income of the lowest income class in 1970; when the original article was written, only the preliminary data from the 1970 *Demographic Census* were available. Using estimates of the average income of the lowest income class now available, the 1970 share would be a smaller 7.5 percent and the rate of growth of absolute income of the poor only 12 percent.[1]

The problem inherent in the simple exclusion of zero-income family workers is apparent in the necessity to redefine the poverty line arbitrarily—increasing it by a third—in order to accommodate the percentages of the labor force approximately consistent with the percentage of families I found to be in poverty in 1960. Fortunately there is a much more satisfactory alternative. That is a direct family income comparison at the two dates, impossible when my original article was written.

II

Admittedly such a comparison cannot be made as accurately as one might like. The 1970 family income data are available only for monetary income, without corrections for income in kind. Comparable 1960 information had been presented in my original article, however. Differences between the distributions over the decade can then be consistently analyzed, subject only to the error of relative changes in the importance of supplementary, nonmonetary income. The distortions introduced by the inclusion of nonpaid family workers are thereby largely avoided: the reported income of household head and other family members partially reflects these unremunerated labor inputs, and very few families fail to report any income at all.

This comparison is preferable in any event because it focuses the poverty concern upon consumption units. The analysis yields conclusions far different than those Fields insists upon. His revisionist findings are completely rejected by the application of his methods to the appropriate family income data set. Absolute gains for the poor are smaller than for the nonpoor, exactly as is true when zero-income workers are excluded.

Table 1 presents the monetary family income distributions for 1960 and 1970, both for the original income ranges. Table 2 converts them to decile form. Zero-income families should be, and are, included. These are families that receive their limited income in kind rather than cash, or temporarily lack earnings and depend upon transfers from others. It is reasonable to infer the vast majority of such families would be classified as poor even were it possible to calculate income in kind or average annual, instead

[1] This recalculation is based upon Carlos Langoni's estimates of mean income by decile for 1970 (p. 64). By interpolation, using a uniform distribution, the mean income for the lowest 13 percent of the economically active income recipients was derived—equivalent to the income class 0–59 NCr$. Note, moreover, that if one uses Langoni's 1960 and 1970 estimates wholly, the results are fully consistent with my calculations here. The share of income of the poor—now 36 percent—declines from 9.6 to 8.4 percent, and the rate of growth of their absolute income is now 19 percent compared to an average gain of 37 percent.

TABLE 1—FAMILY INCOME DISTRIBUTIONS, 1960 AND 1970

Monthly	1960		Monthly	1970	
Income NCr$ 1960	Percent of Families	Percent of Income	Income NCr$ 1970	Percent of Families	Percent of Income
None	2.7	0	None	3.7	0
0–2.1	14.8	2.1	1–50	5.9	.6
2.1–3.3	15.0	4.3	51–100	16.5	3.1
3.3–4.5	12.9	5.5	101–150	13.0	4.1
4.5–6.0	13.1	7.4	151–200	13.9	6.1
6.0–10.0	16.7	14.3	201–250	5.7	3.2
10.0–20.0	15.5	24.2	251–300	7.0	4.8
20.0–50.0	7.5	25.0	301–400	8.0	6.9
50.0±	1.8	17.2	401–500	6.0	6.7
			501–1000	12.1	22.7
			1001–1500	3.6	11.4
			1501–2000	1.9	8.1
			2001±[a]	2.7	22.4
Mean Income 9.2			401		
Gini Coefficient .55			.59		

Source: See my paper, Table 1; *Censo Demografico*, Table 10.

[a]The implicit mean income for the open-ended class was derived from Langoni's mean income for the entire distribution (excluding zero-income families), Table 1.2. Note that this likely understates inequality since it appears Langoni's calculation of open-ended class mean income did not correctly allow for those with incomes above NCr$ 9999; a Pareto extrapolation yields a 31 percent greater estimate for the open-ended class. That would yield a share of 27.3 percent of income for those in the class NCr$ 2001+. Langoni's estimate for the lowest decile was also used to obtain the mean income for the bottom income class. This, too, favors less charge in inequality than application of methods fully comparable to those in 1960.

of monthly, receipts. Moreover, their relative incomes—after adjustment—would remain among the lowest compared to others whose incomes had likewise been recalculated.

It is apparent from the comparisons in the tables that an unequivocal and significant deterioration of the relative family distributions occurred between 1960 and 1970. This is despite estimation methods for 1970 that are biased against greater inequality. The Gini index of concentration increased by almost 10 percent; the position of the lowest 40 percent of families diminished from an income share of 9.4 percent to one of 8.15 percent, a reduction of 14 percent. Poor families fared badly in a relative sense.

But what of absolute incomes and poverty lines? That calculation requires specification not of a single minimum family income, but a set of standards that vary with family size. Small families with limited total income may escape poverty, while larger units with greater incomes fail to satisfy basic needs. The original 1960 estimates were calculated with such a variable poverty line as a function of family size.

Here, because the interest is in changes over a single decade during which demographic trends were unbroken, one can simplify. That continuity is readily verified. Average family size in 1960 was 4.95; in 1970, 4.85. The correlation of family size with family income likewise was stable. The distribution of poverty families by family size in the two years remains similar, when more sophisticated variable poverty lines are applied. Single person families represent 4 percent of the poor in 1960, the same in 1970; those with two and three persons, 18 percent in 1960, 22 percent in 1970; those with four and five persons 27 percent in 1960, 25 percent in 1970; and those with six or more, 51 percent in 1960, 48 percent in 1970.[2]

[2]The 1970 results were obtained by defining a minimum poverty level of NCr$125 monetary income —the minimum rural wage in the Northeast—for an average family size of 4.3; an elasticity of .85 was applied to derive the appropriate minimum levels for

TABLE 2—DECILE DISTRIBUTIONS OF FAMILY INCOME, 1960 AND 1970

1960		1970	
Percent of Families	Percent of Income	Percent of Families	Percent of Income
10−	.5	10−	.65
10	2.2	10	1.7
10	2.9	10	2.4
10	3.8	10	3.3
10	4.9	10	4.3
10	6.0	10	5.5
10	8.1	10	7.5
10	11.0	10	10.6
10	17.0	10	17.9
10+	43.6	10+	46.2

Source: See Table 1. Distributions are obtained by interpolation from Table 1. Each income class was assumed to be uniformly distributed about the mean. This allows for consistent allocation of income, preserving class means, while permitting differential mean income for those allocated to higher and lower deciles.

This stable structure makes possible a simple and direct comparison using the aggregate distributions of Table 1. Family monthly income of 1960 NCr$3.3 may serve as an approximate minimum level of monetary receipts. It is about 10 percent smaller than the rural minimum wage prevailing in the Northeast during the year 1960. Such a boundary includes 32.5 percent of Brazilian families, quite similar to the results of my more sophisticated calculations taking into account income in kind, family size, and regional and rural-urban differences in prices.

Conversion of such a standard to 1970 involves the choice of an appropriate price deflator. As Paul Beckerman and Donald Coes point out in their comment, interpretation is sensitive to which is chosen—a point Fields ignores. At one extreme, the São Paulo cost-of-living deflator implies a 1970 monthly income of NCr$126.3; at the other, the wholesale price index weighted by internal supply yields NCr$106.3. The revised *GDP* implicit deflator implies NCr$114.8, while the cost of living in Guanabara gives NCr$120.4.

The minimum wage prevailing in the rural Northeast at the time of the 1970 census was NCr$125. It corresponds to a family income of about $600 in 1976 dollars, or a per capita income of a little more than $120 in a country boasting of a per capita level of $1,000. Using that wage as the absolute standard is equivalent to recognizing the greater validity of the cost-of-living indexes as a measure of inflation in living standards. Note, moreover, that labor groups have argued that the cost of living of the poor rose differentially to their still greater disadvantage (see *DIEESE*).

Using NCr$125 as the appropriate 1970 poverty line, and one still comparable to 1960, we can infer from Table 1 that 32.5 percent of families—and almost 39 percent of the population—remained in poverty in 1970. That is, there was no change in the relative number of Brazilians falling below a minimum standard of living despite the economic progress of the decade. At best, using the *GDP* deflator to establish a less stringent standard, the relative number of families below the line may have diminished to 29.7 percent, a reduction of 9 percent.

So far, the minimum changes in the number of the poor found even by Fields are confirmed. The much more dramatic differences come in the calculation of income gains of the poor. As has been demonstrated by Montek Ahluwalia et al., Fields' calculations are biased by his incorrect interpolation of the 1970 class limits to conform to

other family sizes. Such a procedure is identical to that used in my earlier article except that adjustments for income in kind, rural-urban and regional differences are precluded (see my paper, pp. 393–94).

the 1960 distribution. Such interpolation introduces another element of indeterminacy in calculations involving the poor and the nonpoor. The problem is how to allocate the income of that income class split by the poverty line to each of the two categories. Because I work with the more finely calibrated 1970 distribution rather than the 1960 distribution, however, that indeterminacy is modest, as is seen below. It is further limited by the reasonable constraint that average income of the poor should be smaller than that of those in the same income class who are assigned to the nonpoor.

As a first approximation, I use a uniform distribution of families within the income class straddling the poverty line, and calculate appropriately different mean incomes corresponding to the two groups. On that basis the average income of families above the line went up by 27 percent over the decade, and those below by only 8 percent, compared to average family income increase of 25 percent.[3] Fields, by contrast, had incomes of the poor growing 63 percent; incomes of the nonpoor, 28 percent; and the average income, 32 percent. It is difficult then to agree with Fields that his novel emphasis upon absolute income growth of the two groups "reinforces the earlier observation that the rich in Brazil did not benefit during the 1960's at the expense of the poor" (p. 575). It is true that absolute decline of family income of the poorest did not occur, but it is equally apparent that their incomes grew only modestly, and less rapidly than those of the rich.

This lower growth of poor family income is invariant to the method of interpolation. At the upper limit, if mean incomes of the poor were almost the same as the nonpoor in the relevant income class, the growth rate of average incomes of the poor would be 12 percent; at the lower limit if their incomes were equal to the class minimum, the rate becomes 4 percent.

The empirical evidence therefore is of a single and compelling piece, whether the criterion be the relative number of poor families or the growth of their average income. The absolute numbers of families below the poverty line increased between 1970 and 1960. Relative to total income, the absolute income transfer required to bring all families up to a minimal standard diminished only marginally between 1960 and 1970. The consequences of intervening real income growth meant that while 5.2 percent of income would have had to be transferred in 1960, a lesser 4.6 percent was required in 1970. Had the government determined seriously to deal with poverty, an enlarged total income would have made the task easier; its weight, however, was largely on the side of bolstering profits.

III

A final brief word is in order about Fields' focus on *income* changes between the poor and nonpoor. The very appeal to income, rather than the relative number of the poor, constitutes a judgment that increased income below the line represents an improvement in welfare. If the conditions of existence of those who are poor remain substantially unchanged despite temporary gains in income, then such an intuitively attractive assumption is no longer valid. Much thus depends upon how the poor react to increases in their income. The thrust of the argument in favor of social provision of basic needs is precisely that small monetary increments to the poor cannot be relied upon to assure adequate nutrition, health, and housing. Poverty gap analyses focusing on income then remain valid in indicating the magnitude of the income transfers required to achieve a minimum standard of living, but not as measures of welfare.

[3] I deflate here and subsequently by the *GDP* deflator for comparability with Fields' calculations. If the larger deflator implicit in the NCr$125 poverty line were used, the differential would be more marked: an average of 15 percent; nonpoor growth of 16 percent; poor decline of .6 percent. If one were to use a differential deflator that was greater for the low-income group, the gap would of course be still more pronounced. Note as well that if the *GDP* deflator is used to define the poverty line, while there is some reduction in number, the rate of growth of income of poor families is still zero. That is because a lower absolute limit also means lower average income of the poor in 1970.

Even if we admit the relevance of income gains to the poor, and thereby accept what Amartya Sen (p. 219) terms the monotonicity axiom, Fields' particular specification is unduly restrictive. It weights income equally within the poor and nonpoor, as well as between them. Are we in fact indifferent to a style of development that benefits the very rich on the one hand, or the middle class on the other? Does each imply equal probability for the eventual reduction of poverty?

Concern for the poor exclusively, moreover, does not condone a procedure that equally weights increases in income whether the beneficiaries are close to or far from the poverty line. The intensity of poverty, and the shortfall of basic needs, is likely to be dependent upon the distribution of the enrichment effect among the poor.

One could do better by employing a Sen poverty index that explicitly takes into account not only increases in income of those in poverty, but also its distribution. His measure is $P=H[I+(1-I)G]$, where H is the percentage of the population below the poverty line; I is the average percentage of the income shortfall of the poor from the minimum level required; and G is the Gini coefficient of the distribution of income among the poor. A higher value of the index indicates worse poverty. This index satisfies not only the monotonicity axiom but also the transfer axiom requiring that a transfer of income from below the standard to anyone above it must increase the measure.

Calculations of P for the 1960 and 1970 family income distributions confirm no significant decrease between the two dates. Indeed, with the exception of H—constant at .325 in both years—the two other components of the index increase when the absolute standard of NCr\$125 is used. The average percentage I rises from .45 to .46. Although the poverty gap normalized on total income falls, the more rapid rise of incomes above the line than below causes the gap to increase slightly as a percentage of the poverty level income.[4] The coefficient G goes from .24 to .26 as the inequality of the income distribution among poor families became greater. The net consequence is a 3 percent increase in the index from .188 to .19. Greater concentration of income among the poor is responsible for about three-fifths of the increase; the larger percentage shortfall from minimum average incomes, the residual. If the lower GDP absolute poverty line is used, P for 1970 is .175 with the entire decline attributable to the reduction in number of poor families. That is exactly contrary, of course, to Fields' insistence upon the importance of enrichment effects.

IV

My conclusion must therefore be that Fields has confused rather than elucidated the complex questions posed by the relationship between income growth and income inequality in Brazil during the 1960's. His revisionism depends largely upon dubious manipulation of the data base, and contributes little to understanding the process of change in Brazilian absolute poverty. No one seriously engaged in the debate regarding the levels and changes in inequality had argued for immizeration of the poor, nor had there been inadequate attention beforehand to the importance of absolute poverty. What had been argued, and what seems to remain true, is that the relative deterioration in income distribution during the decade prevented economic growth from significantly alleviating the burden of crushing absolute poverty that afflicted a third of Brazilian families, and a still larger proportion of the population.

More recent distributional data have become available for 1972, 1974–75, and 1976, and new questions have been raised regarding the direction and timing of changes in Brazilian inequality. Analysis of this information is already underway, with particular interest in the effects of rapid economic growth after 1968. Such continuing consideration and evaluation are to be welcomed,

[4]In this case, the deflator implicit in the NCr\$125 poverty line is applied to incomes of the poor. That explains the apparent inconsistency between the increase in I and the small but positive rate of growth in family income of the poor earlier noted.

particularly when they extend the fragile information base so far utilized. The issue, after all, is of more than academic interest. That is why it is important to give notice of the deficiencies in Fields' original presentation.

REFERENCES

M. Ahluwalia et al., "Who Benefits from Economic Development?: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 242–45.

P. Beckerman and D. Coes, "Who Benefits from Economic Development?: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 246–49.

G. S. Fields, "Who Benefits from Economic Development?—A Reexamination of Brazilian Growth in the 1960's," *Amer. Econ. Rev.*, Sept. 1977, *67*, 570–82.

A. Fishlow, "Brazilian Size Distribution of Income," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 391–402.

Carlos G. Langoni, *Distribuicao da Renda e Desenvolvimento Economico do Brasil*, Rio de Janeiro 1973.

A. Sen, "Poverty: An Ordinal Approach to Measurement," *Econometrica*, Mar. 1976, *44*, 21–31.

Censo Demografico, Brazil 1970; *IBGE* 1973.

Departamento Intersindical de Estatisticases Estudos Socio-Economicso (*DIEESE*), São Paulo, *Estudos Socio–Economicos*, No. 2, Jan. 1974.

Who Benefits from Economic Development?: Reply

By GARY S. FIELDS*

Before the appearance of my 1977 paper in this *Review*, it was widely thought that the income distribution worsened during the economic growth which took place in Brazil during the 1960's. My paper demonstrated that the familiar data, when analyzed from an absolute perspective, could show that the poor *had* benefited from growth. I found that the entire income distribution shifted, benefiting every income class; that the proportion of the economically active population with incomes below the poverty level (as defined by Brazilian standards) declined during the decade; that those who remained poor were less poor than before in absolute terms; and that the rate of growth of income among the poor was at least as great as the rate of growth among the nonpoor. These results came as a surprise to me, and so I did not expect that my conclusion—that Brazil seemed to do better on the income distribution front than many observers had originally thought—would be received uncritically by others.

The preceding comments paint a less rosy picture. Previously used data are shown to be deficient in important respects and new evidence is offered which contradicts the old. Because of this new and more critical evidence, I myself am less certain of what actually happened. But as I shall show, this latest reexamination also confirms some of the more positive aspects of the Brazilian experience. It is fair to say that neither the most favorable nor the most unfavorable position can be sustained unambiguously.

*Associate professor of labor economics and economics, Cornell University. Thanks are due to Paul Beckerman, Albert Fishlow, and Graham Pyatt for discussions of earlier drafts of their comments and of my reply. However, what emerges is my own statement and not necessarily a consensus position on all points.

I

Montek Ahluwalia, John Duloy, Graham Pyatt, and T. N. Srinivasan (hereafter A-D-P-S) criticize my analysis of changing income distribution in Brazil on three counts: limitations of the data used in my study; illogic of an approximation procedure; and qualitatively dissimilar results under seemingly plausible assumptions. Let me deal with these in turn.

We would do well to remember that I used the data from Albert Fishlow's 1972 paper with no adjustment save an interpolation. I used Fishlow's data without modification for methodological reasons—I wanted to show that qualitatively different results would emerge from a different *kind* of analysis (absolute vs. relative). Fishlow's data did not permit the analysis of families nor the use of income group-specific price deflators, both of which A-D-P-S criticize in my work. Nevertheless, some at the World Bank, among them Ahluwalia (1974), have in the past regarded Fishlow's data as reliable enough to cite conclusions derived therefrom and to use the data in their own research.

Ahluwalia et al.'s second point concerns my interpolation assumptions. I assumed that all those in the income class NCr$0–2.8 had the same income, that 75 percent of them (=2.1, the poverty line ÷2.8, the range of the income bracket) were "poor," and that the remaining 25 percent of them were not. Under these assumptions, I estimated the proportion poor as 35.5 percent of the population in 1970. Ahluwalia et al. regard these assumptions as illogical, and in at least one way they are right: if the poor are defined as those with incomes below NCr$2.1, the other 25 percent in the income group 0–2.8 should properly have been considered poor as well, since by assumption,

their incomes were below the poverty line also. By referring to the poorest 35.5 percent as the poor, a misstatement arose. In footnote 6 of my 1977 paper, as quoted in A-D-P-S's introductory paragraph, I stated that the average income of the poor must have risen at an above average rate; that statement should have been expressed conditional on 35.5 percent being the proportion poor. Ahluwalia et al. accurately noted my imprecision. One implication of their critique is that the number 35.5 percent merits less weight than I gave it in the 1977 paper. This point has a bearing on the interpretation of other of their results, discussed below.

Ahluwalia et al. are quite correct on their third point: qualitatively dissimilar results *are* possible based on Fishlow's 1972 data. They have ingeniously displayed the various possibilities for the proportion poor (i.e., incomes less than NCr\$2.1) and their income share. I would prefer to state their conclusions in the following way:

1) It is logically possible, though not proven, that: There was a greater percentage poor in Brazil in 1970 than in 1960.

2) It is logically possible, though not proven, that: The average income of the poorest 35.5 percent grew at a lower rate than the incomes of the other 64.5 percent between 1960 and 1970.

3) Statements 1) and 2) cannot both be true, that is: It is true either that the percentage poor fell or the average income of the poorest 35.5 percent grew faster than the average for the nonpoor or both.

4) With certainty, we can say from Fishlow's original data that: The average income of the poorest 37.0 percent (the proportion in the two lowest income classes in 1960) grew at an above average rate.

What are we to make of these various findings? Following the logic of A-D-P-S's argument, the figure 35.5 percent poor (my estimate for 1970) should be accorded no particular significance, since it was derived under assumptions which A-D-P-S regard as inappropriate. We have all accepted 37.0 percent as the proportion poor in 1960, though even that figure is somewhat arbitrary. Suppose that 37.0 percent is used as a reference figure for 1970. How did the incomes of the poorest 37.0 percent change over the decade? From Fishlow's 1972 data, I can only conclude that they benefited from growth at an above average rate, as the following paragraph shows.

How much above average was the growth rate of income of the poorest 37.0 percent? The answer depends on the particular assumption made concerning the shape of the income distribution within the NCr\$0–2.8 income class. Various possibilities are illustrated in my 1976 paper. There, I showed that the poorest 37.0 percent of the population received *at least* 5.4 percent of the income in 1970, which is the same as A-D-P-S's lower limit (equation (2′) evaluated at $P=37.0$ percent).[1] Since the poorest 37.0 percent received a smaller share, 5.2 percent, in 1960, their share *rose* even assuming the *minimum* possible increase. And this minimum increase is based on perfect *inequality* of income distribution among those earning less than NCr\$2.8, hardly a plausible assumption.[2] The more equal the distribution within the 0–2.8 group, the *greater* the growth in income share of the poorest. Hence, under any consistent assumptions the average income of the poorest 37.0 percent grew at an above average rate.

In short, my main empirical conclusion —that the poor in Brazil experienced percentage income gains at least as great as those of the nonpoor—holds up to the A-D-P-S critique, if in both years we define the reference group of the poor as 37.0 percent (which is the unquestioned figure based on the 1960 census) and if we accept Fishlow's original data for both years. However, as A-D-P-S rightly point out, if we seek to interpolate the number poor and their average income, a wide range of possibilities is consistent with the published data. It is not *certain* from these data that a

[1]I report 5.49 percent in my 1976 paper. Ahluwalia et al.'s constraint implies a value of 5.41 percent. The difference is due to rounding.

[2]Perfect inequality is where some in that income class earn NCr\$2.8 and the rest earn zero, the number of each being determined by the amount of total income which that group receives.

smaller proportion of the population fell below a constant real absolute poverty line and that the average income among those remaining poor increased, though this *may* have been the case. Less favorable outcomes are also consistent with the available data. But it remains *impossible*, given our acceptance of the original data, that everything went wrong in the sense of both a growing proportion in poverty and a decline in their average income relative to that of the rest of the population. Of course, if the original data are regarded as unsuitable for poverty analysis, anything is possible, as the other commentators seek to demonstrate.

II

Paul Beckerman and Donald Coes (hereafter B-C) have recalculated the absolute poverty estimates in my paper by replacing the implicit *GDP* price deflator which I used by the price index for São Paulo. The two cost-of-living indices are 35.32 (implicit deflator) and 38.26 (São Paulo index). The difference between these numbers does not seem great, nor does the difference between the annual rates: 42.8 and 44.0 percent. Yet, by using the São Paulo index, estimated absolute poverty in 1970 is much greater and the participation of the poor in economic growth is correspondingly reduced.

I differ with the authors' dismissal of the absolute poverty approach to income distribution change, specifically their claim that "... the method itself cannot provide meaningful quantitative statements about changes in income earned by different groups, due to its sensitivity to small changes in the measured rate of real income growth" (p. 249). They reach this conclusion based on the judgment that the differences in price index are "small."

In fact, the differences are large. Mean income went from NCr$5.52 in 1960 to NCr$258.1 in 1970, both expressed in current cruzeiros. With the price index I used, mean income in constant cruzeiros comes to 7.31. With the index they suggest, the mean is 6.74. By my estimates, mean income increased by 32 percent over the decade. By their estimates, the increase in mean income was 22 percent. Thus, the overall rate of growth is reduced by *one-third* if the São Paulo price index is used rather than the implicit price deflator! This differential is hardly small. Think what it does to the whole macro-economic discussion of Brazilian growth, especially the post-1967 "miracle."

Compared with this macro-economic effect, the consequences of using a different price index seem rather small, for example, their estimate of percentage poor in 1970 as 37.3 vs. mine of 35.5 percent. Note too that by their estimates the incomes of the poor grew faster than those of the nonpoor (38 vs. 22 percent). Qualitatively, at least, my published results hold up.

The conclusions I would draw from B-C's findings differ from theirs. I would conclude that macro-economic data on real income change are themselves highly sensitive to the price index used; that estimates of change in absolute income and absolute poverty are similarly sensitive; that their results reaffirm my conclusion that the poor did share at least proportionally in the economic growth of Brazil; and that relative inequality indices being totally insensitive to all this are not very helpful.[3] These are all matters of interpretation where we might disagree.

On one issue, B-C are in error. They state in their concluding paragraph: "The more traditional focus on relative inequality reflects the theoretically and empirically plausible assumption that the marginal utility of income is decreasing, while the absolute shares approach deliberately ignores this motive...." The second part of that statement is simply wrong. The only reason why anyone would want to calculate changes in absolute poverty is to focus in on those who presumably have the highest marginal utility of income (the poor) to the exclusion of those whose marginal utilities are assumed to be lower.

III

Albert Fishlow's comment presents new evidence claiming that poverty did not diminish in Brazil in the 1970's. My 1977

[3]Obviously, their Lorenz curve is the same as mine.

paper claimed that it did. What is not at issue is the reliability of the source of the basic data, since in both instances, the source is Fishlow himself (or more precisely, his rendering of Brazilian census data).[4] What *is* at issue is which data are most appropriate.

Fishlow bases his claim that poverty did not diminish on a data set different from the one he used earlier. His 1972 paper examined changes in income inequality between 1960 and 1970 among *individuals* in the economically active population. His present note, on the other hand, uses data released in the interim to compare *families*. If the Brazilian data are to be believed, what they are telling us is that poverty was lessened among individuals but not among families. Thus, an assessment of the distributional performance of the Brazilian economy in the 1960's turns on the choice of recipient unit. Fishlow claims that family income comparisons are clearly superior for studying income distribution change. I disagree.

Other, more specific, points are raised in Fishlow's comment. He writes that my manner of constructing a poverty line is "both critical to the findings and quite illegitimate" (p. 250). He is wrong on both counts. Drawing the poverty line at 2.1 is not critical. The same qualitative results would have been found had the poverty line been drawn at 3.3 (the cutoff of the next higher income bracket). In the absence of reason to suppose otherwise, it is certainly legitimate and defensible to do what I did: to approximate the proportion of individuals who are poor by the proportion of families which are poor. He complains that I used uncorrected data. That is legitimate when only uncorrected data are available. In his 1972 paper, Fishlow did *not* correct for income in kind or for regional price differences *in making intertemporal income distribution comparisons*, and so neither did I when I used his data. It is legitimate, I submit, to include individual zero-income recipients, since unemployment reduction is an important means of poverty alleviation.[5] To exclude zero-income workers is, I think, a cure worse than the disease. But I must take responsibility for an unpublished arithmetic error which invalidates the statement in footnote 7 of my 1977 paper—the results *are* materially affected if zero-income workers are excluded.

Some of Fishlow's calculations are mutually inconsistent. His note argues that the correct poverty line for families is NCr$3.3 for 1960 and an equivalent poverty line for 1970 is NCr$125. Even if we accept these figures as equivalent, and I hesitate to, the ratio of prices is $125/3.3=37.9$. This implies that the 1960 mean income (NCr$9.2) is equivalent to NCr$348.5 when measured in 1970 prices in nominal terms. The mean income in 1970 was NCr$401. The rate of growth of mean income is therefore $(401-348.5)/348.5-1=15$ percent, not 25 percent as Fishlow reports. Either the appropriate poverty line is not NCr$125 or the rate of growth is not 25 percent. Fishlow cannot have it both ways. Which way he has it considerably affects his comparisons of income growth of the poor relative to the average. And let me record my hesitation in accepting Fishlow's calculation of changes in the Sen index based on the type of information available, especially since there is an inconsistency between the claim that the *I* component of the Sen index increased and the earlier observation that the poor's income rose albeit slightly.

[4]Fishlow's tabulations go beyond the published data in at least one important way. The census publications do not report actual income shares for each income group. Fishlow published derived shares, which he obtained using a two-way procedure: in the case of the lowest income bracket, by fitting a Pareto distribution; for the other income brackets, by setting the mean equal to the midpoint. This procedure introduces particular assumptions about the distribution of income *within* the group classified as poor. We have no way of knowing whether these assumptions are or are not accurate. When I wrote my 1977 paper, I did not know that the income shares published by Fishlow were fitted, not actual, nor apparently did the other commentators in preparing their comments.

[5]The census evidence cited by Fishlow showing that most of the unremunerated workers were found in agriculture, not in unemployment, is in direct conflict with the findings of P. I. Singer on which I had based my earlier judgment.

The new evidence presented in Fishlow's comment offers us a choice—on which income recipient unit (families or individuals) is most appropriate for purposes of intertemporal analysis, on whether a concern with income or expenditure distribution is of more interest, on the legitimacy or illegitimacy of uncorrected income distribution data. I believe I might be forgiven if in 1975 (when the first draft of my paper was written) or in 1976 (when the final draft was accepted for publication), I accepted Fishlow's earlier decisions on these questions, especially when it is recognized that I was trying to show that his own data, analyzed with a different *type* of measure, would suggest a quite different interpretation. I would conclude that Fishlow has *not* presented a "single and compelling" piece of evidence to the effect that the poor in Brazil did not share in economic growth. Maybe they did, maybe they didn't, but his results do not sustain an unambiguous conclusion either way.

IV

My 1977 paper had two purposes, one methodological and one empirical. The methodological goal was to apply in the case of a less-developed country a largely overlooked class of absolute measures which gauge directly the extent to which the poor gain from economic development. At the time I wrote the paper (1975), absolute poverty measures had seldom been used in a *dynamic* context in studies of *LDC*'s (i.e., to measure who benefits how much from economic development within a country, though Fishlow had effectively introduced these measures to construct static poverty *profiles* and Ahluwalia had used these measures in a *cross section* of countries). I hoped to show that those who wish to give predominant weight to countries' progress toward alleviating economic misery might find these absolute measures (changes in proportion of income units which are poor and changes in average income among the poor) more convenient than the more familiar measures of relative inequality (changes in Lorenz curves, Gini coefficients, income shares of particular percentile groups, etc.) for gauging the beneficiaries of growth.

The empirical goal was to reexamine the specific case of Brazilian growth in the 1960's. Toward that end, my results established that the same income distribution data, when analyzed from an absolute rather than from a relative perspective, yielded a qualitatively distinct and decidedly more positive picture of who benefited.

I believe the methodological objective has been largely satisfied. Though my paper provoked much discussion, pro and con, I am unaware that anyone on either side has rejected *in principle* the call for applying absolute tools to the study of income distribution change, at least in conjunction with relative inequality measures if not as a replacement for them, though Beckerman and Coes reject absolute measures *in practice*. We are not likely to witness a return to the debates of the mid-1970's over whether the participation of the poor in economic growth is better measured by the Gini coefficient rather than by a Theil index, Kuznets ratio, Atkinson index, Pareto coefficient, *log* variance, or what have you. To the contrary, the development community is now groping toward the most appropriate way of measuring the alleviation of absolute poverty. This concern is reflected in such current phrases as "redistribution with growth," "meeting basic needs," "new directions in development assistance," "progress and commitment for the poor majority," "trickle down," and "distributional weights."

The empirical issue remains unsettled. In the last two or three years, the absolute approach has been applied to the study of distribution and development in other countries besides Brazil. The existing literature is surveyed and new evidence presented in my forthcoming book.

Scholars of integrity welcome the opportunity to subject earlier ideas to tests on new and better data. The evidence presented in the three comments, though new, is not necessarily better. But it is disturbing. The additional evidence shows recent Brazilian economic history in a less favorable light

than I portrayed it before, though in my judgment, less dismal than Fishlow's comment suggests. This can only raise new doubts on the extent to which the poor shared in Brazilian growth.

The several comments have also raised important questions about the suitability of published income distribution data for absolute poverty analysis. Since the appearance of my 1977 paper, and indeed in response to it, the Brazilian data base has come under close scrutiny and some technical limitations brought to attention. In addition to the fine points about the specifics of census reporting procedures, and tabulations derived therefrom (see fn. 4), doubts about the quality of the underlying data are also raised by the very observation that, given a nearly constant demographic structure, the family and individual distributions produce such divergent results.

Attention is rarely given to technical matters such as these; too often, we simply accept whatever data are available. But as the new evidence indicates, these technical issues are paramount in coming to even a qualitative judgment on distributional aspects of Brazilian development.[6]

Notwithstanding our differences, I expect that the several commentators would join me in two final observations: given the available resources, much more could have been done than was done to alleviate economic misery in Brazil; and much more can be done in the future if the political will is there.

[6]It might be noted that technical considerations were also prominent in earlier debates on changing income distribution in the course of the economic development of India.

REFERENCES

M. Ahluwalia, "Income Inequality: Some Dimensions of the Problem," in Hollis B. Chenery et al., eds., *Redistribution with Growth*, New York 1974, 3–37.

______ **et al.,** "Who Benefits from Economic Development?: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 242–45.

P. Beckerman and D. Coes, "Who Benefits from Economic Development?: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 246–49.

G. S. Fields, "More on Changing Income Distribution and Economic Development in Brazil," disc. paper no. 244, Econ. Growth Center, Yale Univ., Apr. 1976.

______, "Who Benefits from Economic Development?—A Reexamination of Brazilian Growth in the 1960's," *Amer. Econ. Rev.*, Sept. 1977, *67*, 570–82.

______, *Poverty, Inequality, and Development*, Cambridge 1980.

A. Fishlow, "Brazilian Size Distribution of Income," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 391–402.

______, "Who Benefits from Economic Development?: Comment," *Amer. Econ. Rev.*, Mar. 1980, *70*, 250–56.

NOTES

The ninety-third annual meeting of the American Economic Association will be held in Denver, Colorado, September 5–7, 1980. The Professional Placement Service will be held December 28-30, 1980, at the Dallas Hilton, Dallas, Texas. Please note that there will be no employment service at the September meetings. Complete details will be published in the June 1980 issue of this *Review*.

The Secretary of the American Economic Association wishes to announce the results of the mail balloting for officers which took place during the fall of 1979. The following individuals took office during January 1, 1980: President-Elect William J. Baumol; Vice Presidents Carl F. Christ and H. Gregg Lewis; and Executive Committee members Martin S. Feldstein and Robert E. Lucas, Jr.

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada, that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to air fare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for applications to be received in the ACLS office are: meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 345 East 46th St, New York, NY 10017, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting. Even when plans are incomplete, a prospective applicant should request forms in advance of the cut-off date, since deadlines are firm and no exceptions are permitted. Awards will be announced approximately two months after each deadline.

The Division of Policy Research and Analysis of the National Science Foundation will continue to support studies of science and technology policy related to: innovation processes and their management; socioeconomic effects of science and technology; environment, energy, and resources; and technology assessment and risk analysis. Proposals may be submitted at any time, but a period of approximately six months is required for review and notification of the award decision. Further information may be obtained by requesting the *Program Announcement for Extramural Research* (NSF 78-78) from the Division of Policy Research and Analysis, National Science Foundation, Room 1233, 1800 G Street, NW, Washington, DC 20550.

The Public Utilities Forecasting Conference will be held at Nottingham University, England, March 25–29, 1980. It is an international meeting and will feature both invited and contributed papers. The proceedings are to be published.

The first annual Sewanee Economics Symposium will be held April 3–5, 1980: "Business in the 'New South': A Historical Perspective." There will be panels on the early problems, modern success, and future possibilities of southern business and industry, and the record of its development. For further information, contact Professor Marvin E. Goodstein, Department of Economics, The University of the South, Sewanee, TN 37375.

The second annual Middlebury College Conference on Economic Issues, "Welfare Reform: Goals and Realities," will be held April 17–19, 1980. The objectives are: evaluation of the antipoverty effort so far; discussion of welfare reform alternatives; and prediction of how new program initiatives would change work behavior, productivity, and marital stability.

The College Curriculum Support Project of the Bureau of the Census is preparing a multidisciplinary student workbook and manual of classroom exercises and projects which use Census Bureau data to analyze real world problems. The workbook will be designed to be used in geography, business, sociology, marketing, statistics, economics, political science, and psychology classes. It will help students understand the uses and limitations of census data, relate concepts in disciplines to contemporary issues, develop critical skills for analyzing reports, and learn how to use survey data when census sources are out of date. Information is needed from census users to assist in the development of these materials: instructors who have developed

student exercises, workbooks, case studies, or manuals are encouraged to submit these materials for possible inclusion in the publication; and practitioners are encouraged to submit examples (or bibliographic citations) of business and governmental studies which have used any of the census-related skills mentioned above. Please send material immediately, or contact Dr. Les Solomon, College Curriculum Support Project, Data User Services Division, Bureau of the Census, Washington, DC 20233. The workbook will be published by the Bureau of the Census and should be ready for classroom use by spring 1981.

Applications are invited for five postdoctoral fellowships available to persons interested in preparing themselves for positions in mental health policy analysis and program planning. Applicants should hold a doctorate in a relevant discipline: anthropology, economics, education, law, management, pediatrics, political science, psychiatry, psychology, or sociology. The program involves one year of study and includes a field placement. Stipends range from $10,000 to $14,000, and the application deadline is March 30, 1980. Notification of awards will be made on May 1, 1980, with the training program to begin on July 1, 1980. Applicants should supply complete biographical information, a statement detailing interests in policy analysis and future career plans, and the names of at least six references. Address applications to Mental Health Policy Program, Vanderbilt Institute for Public Policy Studies, Box 1516, Station B, Nashville, TN 37235.

Seventh Regional Science Dissertation Competition: Three prizes of $700 each will be awarded to the best dissertations in regional science. The studies will be judged on the basis of significance for public policy, significance of the contribution to the analysis of development issues, imaginativeness, timeliness of subject matter, and technical competence. The competition is open to any graduate student whose dissertation has been accepted in partial fulfillment of a doctoral degree in a U.S. university during the period July 1, 1979 and June 30, 1980. Previous years' entrants have come from students in economics, planning, geography, regional science, and civil engineering. Further details may be obtained from Geoffrey J. D. Hewings, Secretary, Regional Science Association, 220 Davenport Hall, University of Illinois, Urbana, IL 61801. The closing date for entries is August 15, 1980.

Call for Papers: The Journal of Family Issues plans to publish a special issue on "Dual-Worker Families" in June 1981 to contain such subjects as: family roles and organization in dual-earner families; their relationships with kin, the economy, and other extra-familial groups and institutions; trends in their incidence, divorce rates, and other indicators of family dynamics; and varying types of dual-earner families (commuter families, etc.). Scholars are encouraged to submit manuscripts no later than September 1980 to Joan Aldous, Department of Sociology and Anthropology, University of Notre Dame, Notre Dame, IN 45556.

New Journal: Call for Papers—The Journal of Labor Research will be issued semi-annually in 1981, and quarterly thereafter. Its objective is to enhance understanding of important issues involving labor unions as organizations by stimulating and renewing interest in such questions as the goals of labor unions; why and how these goals are pursued and achieved. Papers relating to labor economics, labor relations, the political objectives of unions, international as well as domestic labor topics, interdisciplinary studies, and research on public policy issues are of particular interest. All manuscripts should be sent to *Journal of Labor Research*, Department of Economics, George Mason University, Fairfax, VA 22030.

Deaths

Helen A. Cameron, professor emeritus, Ohio State University, Oct. 5, 1979.

William M. Dickson, associate professor, Simon Fraser University, Nov. 1, 1978.

Retirements

James G. Allgood, associate professor emeritus, department of economics and business, North Carolina State University, June 30, 1979.

Charles Hoffman, State University of New York-Stony Brook: dean of social sciences and professor of economics, Queens College, City University of New York, Sept. 1, 1979.

Edwin A. Proctor, professor emeritus, department of economics and business, North Carolina University, Dec. 31, 1978.

Craig D. Woodruff, professor of management, American Graduate School of International Management, Dec. 31, 1978.

Visiting Foreign Scholars

Peter Boumberger, Swiss National Bank: visiting scholar, Graduate School of Business, University of Chicago, Sept. 1979.

Charles Harvey, University of Sussex, England: visiting lecturer in economics, Williams College, July 1, 1979.

Muhammad A. Malallah, University of Jordan: visiting assistant professor of finance, world business department, American Graduate School of International Management, May 31, 1979.

Pavle Sicherl, University of Ljubljana, Yugoslavia: visiting lecturer in economics, Williams College, Feb.–May 1980.

Promotions

Bernard E. Anderson: professor of industry, Wharton School, University of Pennsylvania, 1979.

Saul Z. Barr: executive director, Council on Economic Education in Maryland, Towson State University, Aug. 1979.

Kwang-wen Chu: professor of economics, California State University-Fullerton, Sept. 1, 1979.

R. L. Clark: associate professor, department of economics and business, North Carolina State University, July 1, 1979.

Robert Frank: associate professor of economics, Cornell University, Oct. 1978.

K. Celeste Gaspari: assistant professor of economics, Wellesley College, Aug. 1979.

James Grant: assistant professor of economics, Wellesley College, Aug. 1979.

D. N. Hyman: professor, department of economics and business, North Carolina State University, July 1, 1979.

M. P. Loeb: associate professor, department of economics and business, North Carolina State University, July 1, 1979.

Howard P. Marvel: associate professor of economics, Ohio State University.

Richard A. Palfin: associate professor of economics, Northern State College, Aug. 1979.

Uri Possen: associate professor of economics, Cornell University, June 1, 1979.

Barbara L. Reed: assistant professor of accounting, American Graduate School of International Management, Aug. 22, 1979.

Irwin D. Sandberg: vice president, Open Market Operations and Treasury Issues Function, Federal Reserve Bank of New York, Oct. 1, 1979.

Richard E. Schuler: associate professor of economics, Cornell University, Oct. 1978.

Thomas A. Wolf: associate professor of economics, Ohio State University.

Administrative Appointments

Roger Bolton: chairman, department of economics, Williams College, July 1, 1979.

Gordon L. Brady, Brookings Institution: chief, Economic Analysis Division, National Commission of Air Quality, Washington, D.C., May 9, 1979.

Meredith O. Clement: chairman, department of economics, Dartmouth College, July 1, 1978.

Irwin Feller: director, Institute for Policy Research and Evaluation, Pennsylvania State University, Nov. 1, 1977.

Mukul K. Majumdar: professor and chairman, department of economics, Cornell University, July 1, 1978.

John M. Munro: vice president, academic, Simon Fraser University, June 1, 1979.

Carl J. Palash: chief, research support department, Research Processing Division, Federal Reserve Bank of New York, July 19, 1979.

Les Seplaki: chair, economics department, Rutgers University-Camden, Nov. 1, 1979.

Gordon W. Smith: chairman, department of economics, Rice University, July 1979.

Lawrence H. Thompson, U.S. Department of Health, Education, and Welfare: associate commissioner for policy, Social Security Administration, Jan. 1979.

Paul N. Van de Water, U.S. Department of Health, Education, and Welfare: director, Office of Policy Analysis, Social Security Administration, Jan. 1979.

Appointments

Michael D. Andrews: economist, international research department, Developing Economies Division, Federal Reserve Bank of New York, Aug. 29, 1979.

Dale G. Bails: assistant professor of economics, Memphis State University, Sept. 1979.

Fred Bateman, Indiana University: professor, department of economics, University of the South, Jan.–May 1980.

D. L. Baumer: department of economics and business, North Carolina State University, Aug. 20, 1979.

Robert A. Blewett: assistant professor, department of economics, Western Illinois University, Aug. 27, 1979.

M. E. Bond: professor of economics, Memphis State University, July 1979.

E. H. Brooks: lecturer, department of economics and business, North Carolina University, Aug. 20, 1979.

John L. Bungum: associate professor of economics, Gustavus Adolphus College, Fall 1979.

John A. Burghardt: research economist, Mathematica Policy Research, Inc., Oct. 1, 1979.

Richard C. Cahaan, Cornell University: lecturer, department of economics, Ohio State University.

John F. Chant, Carleton University: professor of economics, Simon Fraser University, Sept. 1, 1979.

Daniel Christiansen: assistant professor of economics and director, Center for Economic Education, University of Northern Iowa, Aug. 1979.

Timothy W. Cooke, Johns Hopkins University: assistant professor of economics, Rice University, July 1979.

Paul R. Cullinan, Syracuse University: economist, Office of Policy Analysis, Social Security Administration, Oct. 1979.

Christine M. Cumming: economist, international research department, Industrial Economies Division, Federal Reserve Bank of New York, Sept. 10, 1979.

Roger A. Dahlgran: assistant professor, department of economics, Iowa State University, Dec. 1, 1979.

Robert H. De Fina: economist, domestic research department, Business Conditions Division, Federal Reserve Bank of New York, Sept. 17, 1979.

James N. Dertouzos, Stanford University: associate economist, economics department, Rand Corporation, Sept. 1979.

Arthur S. De Vany, University of Chicago: professor of economics, Simon Fraser University, Sept. 1, 1979.

Woodrow F. Dick, Jr.: instructor, department of economics, Iowa State University, Sept. 1, 1979.

Dennis D. DiPietre: instructor, department of economics, Iowa State University, Sept. 1, 1979.

David Easley: assistant professor, department of economics, Cornell University, Sept. 1979.

Gerald Epstein, Princeton University: assistant professor of economics, Williams College, July 1, 1979.

E. A. Estes: assistant professor, department of economics and business, North Carolina State University, Aug. 20, 1979.

Frank A. Fernandez: economist, international research department, Developing Economics Division, Federal Reserve Bank of New York, Aug. 1, 1979.

John W. Fuller, National Transportation Policy Study Commission and Wisconsin Department of Transportation: director, Institute of Urban and Regional Research, and professor, departments of economics and geography, University of Iowa, Sept. 1979.

Laurie S. Goodman: economist, international research department, Balance of Payments Division, Federal Reserve Bank of New York, Aug. 29, 1979.

Thomas W. Grannemann: research economist, Mathematica Policy Research, Inc., Sept. 12, 1979.

Kenneth Guetner: economist, domestic research department, Financial Markets Division, Federal Reserve Bank of New York, June 18, 1979.

Aaron S. Gurwitz: association economist, Rand corporation, Washington, D.C.; also economics department, Santa Monica, Feb. 1979.

George Hay: professor of economics and law, Cornell University, Sept. 1, 1979.

W. John Heaney: assistant professor of financial theory, Simon Fraser University, June 8, 1979.

Terence M. Heaps: instructor of natural resources, Simon Fraser University, Sept. 1, 1979.

R. Bryce Hool, University of Wisconsin-Madison: associate professor, department of economics, State University of New York-Stony Brook, Sept. 1, 1979.

Jerry L. Ingles: assistant professor, department of economics, University of the South, Sept. 1, 1978.

Walter Isard: professor, department of economics, Cornell University, Feb. 1, 1979.

J. Thomas Janz, University of Waterloo: assistant professor of organizational behavior, Simon Fraser University, Sept. 1, 1979.

Richard A. Jensen, Northwestern University: lecturer, department of economics, Ohio State University.

Raghbendra Jha, Columbia University: assistant professor of economics, Williams College, July 1, 1979.

Leland L. Johnson, U.S. Department of Commerce: senior economist, Rand Corporation, Washington, D.C.; also economics department, Santa Monica, Sept. 1979.

Roselyne Joyeux: assistant professor, department of economics, Cornell University, Sept. 1979.

Albert Keidel III, National University of Tokyo: assistant professor, department of economics, Ohio State University.

Harry B. Keller: professor of management, American Graduate School of International Management, Aug. 23, 1979.

Kenneth W. Kendall, Dalhousie University: assistant professor of marketing, Simon Fraser University, Sept. 1, 1979.

Daphne Kenyon: instructor, department of economics, Dartmouth College, July 1, 1979.

Meir Kohn: associate professor, department of economics, Dartmouth College, July 1, 1979.

M. W. Kusnic: assistant professor, department of economics and business, North Carolina State University, Aug. 20, 1979.

Loretta L. Landy: economist, international research department, Industrial Economies Division, Federal Reserve Bank of New York, July 11, 1979.

Young Goo Lee: assistant professor, department of economics, State University of New York-Stony Brook, Sept. 1, 1979.

Arleen Leibowitz: economist, economics department, Rand Corporation, Sept. 1979.

Charles Lieberman: economist, domestic research department, Monetary Analysis Division, Federal Reserve Bank of New York, Sept. 5, 1979.

Thomas S. McCaleb, University of Kansas: assistant professor of economics, Rice University, July 1, 1979.

John C. P. McCallum, University of Manitoba: assistant professor of economics, Simon Fraser University, Sept. 1, 1979.

J. Huston McCulloch, Boston College: associate professor, department of economics, Ohio State University.

James N. McGowen: instructor, department of economics, University of the South, Sept. 1, 1979.

Robert McGuire, Ball State University: visiting assistant professor, department of economics, Ohio State University.

Joseph S. Mascia: assistant professor, department of banking, economics, and finance, School of Banking and Money Management, Adelphi University, Sept. 1, 1979.

Steven A. Meyer: assistant professor, economics department, College of William and Mary, Sept. 1979.

William H. Meyers: assistant professor, department of economics, Iowa State University, Aug. 27, 1979.

Cathy G. Miners: economist, domestic research department, Financial Markets Division, Federal Reserve Bank of New York, Sept. 10, 1979.

Dennis L. Nef: instructor, department of economics, Iowa State University, Sept. 1, 1979.

Henry H. Otterman: instructor of management, American Graduate School of International Management, Aug. 28, 1978.

R. B. Palmquist: assistant professor, department of economics and business, North Carolina State University, Aug. 20, 1979.

D. Terry Paul: visiting assistant professor, department of economics, Iowa State University, Sept. 1, 1979.

Vaman Rao: associate professor, department of economics, Western Illinois University, Aug. 27, 1979.

David L. Roberts: economist, international research department, Developing Economies Division, Federal Reserve Bank of New York, Sept. 12, 1979.

Kathy A. Ruffing, U.S. Department of Labor: economist, Office of Policy Analysis, Social Security Administration, Oct. 1979.

C. D. Safley: assistant professor, department of economics and business, North Carolina State University, Jan. 1, 1980.

Beverly A. Spikes: instructor of accounting, Simon Fraser University, Sept. 1, 1979.

Peter J. E. Stan, Harvard University: associate economist, economics department, Rand Corporation, Sept. 1979.

Steven Swidler, Brown University: visiting assistant professor of economics, Rice University, 1979-80.

Dean Tjosvold, Pennsylvania State University: associate professor of organizational behavior, Simon Fraser University, Sept. 1, 1979.

Barbara H. Tuckman: assistant professor of economics, Memphis State University, Sept. 1979.

Howard P. Tuckman: professor of economics, Memphis State University, Sept. 1979.

Ramakrishna Vaitheswaran: visiting associate professor, department of economics, Iowa State University, Sept. 1, 1979.

James H. Vanderhoof, University of North Carolina-Chapel Hill: lecturer, department of economics, Ohio State University.

Paul A. Wachtel: economist, domestic research department, Business Conditions Division, Federal Reserve Bank of New York, Sept. 12, 1979.

William L. Wilby: visiting instructor of economics and finance, American Graduate School of International Management, Jan. 25, 1979.

Ralph A. Winter, University of California-Berkeley: assistant professor of economics, University of Toronto, July 1, 1979.

George Zodrow, Princeton University: assistant professor, department of economics, Rice University, July 1979.

Leaves for Special Appointment

Bernard E. Anderson, University of Pennsylvania: director, Social Science Division, Rockefeller Foundation.

Wayne Archer, University of Florida: Office of Economic Research, Federal Home Loan Bank Board, Washington, D.C., 1979-80.

M. Sabry El-Shabrawy, American Graduate School of International Management: associate professor of management, American University, Cairo, 1979-81.

Gillian Garcia, University of California-Berkeley: financial economist, Division of Economic Research and Analysis, Office of the Comptroller of the Currency, U.S. Treasury Department, 1979-80.

Roy J. Gardner, Iowa State University: National Science Foundation-National Center for Scientific Research Exchange Scientists, France, Aug. 1, 1979-June 30, 1980.

Cecil G. Gouke, Ohio State University: senior Fulbright lecturer, University of Dar es Salaam, Tanzania, 1979-80.

Michael J. Hamburger, Federal Reserve Bank of New York: visiting professor of economics, School of Business Administration, New York University, Oct. 1979–Sept. 1980.

Glenn V. Henderson, Jr., Louisiana Tech University: visiting professor, University of Hawaii-Hilo, 1979-80.

Whitney Hicks, University of Missouri-Columbia: Fulbright professor, University of Nuevo Leon, Monterrey, Mexico, 1979-80.

Christopher I. Higgins, Minister, Economic and Financial Affairs, Canberra: Australian delegation to O.E.C.D., Paris, Jan. 1980–Dec. 1982.

Estelle James, State University of New York-Stony Brook: visiting scholar, Tel-Aviv University, Jan. 1–June 1, 1980.

Charles Orvis, Southwestern University at Memphis: economist, Office of the Assistant Secretary for Policy and International Affairs, U.S. Department of Transportation, 1979-80.

Mary Rieder, Winona State University: senior economist, Division of Medicaid Cost Estimates, Health Care Financing Administration, U.S. Department of Health, Education, and Welfare, 1979-80.

Steven H. Sandell, Ohio State University: Brookings Institution.

Michael L. Visscher, Ohio State University: Duke University.

Betsy B. White, Federal Reserve Bank of New York: presidential executive interchange program, Chase Manhattan Bank, Oct. 1979-Sept. 1980.

Resignations

D. L. Brito, Ohio State University: Tulane University, June 1979.

Donald Gordon, Simon Fraser University, Dec. 31, 1978.

Daniel Granot, Simon Fraser University, Aug. 31, 1978.

Charles E. McLure, Jr., Rice University: vice president, National Bureau of Economic Research, June 1979.

William H. Oakland, Ohio State University: Tulane University, Sept. 1979.

John G. Pomery, Rice University: Northwestern University, June 1979.

Samuel Rosenberg, Williams College, June 30, 1979.

Roy J. Rotheim, Franklin and Marshall College: executive editor, *Challenge*, July 1, 1979.

Herman O. Stekler, State University of New York-Stony Brook: Institute for Defense Analysis, Washington, D.C., Sept. 1, 1979.

Michael D. Yokell, Solar Energy Research Institute: president, Resource Management Consultants, Inc., Boulder, CO.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1-Deaths
2-Retirements
3-Foreign Scholars (visiting the USA or Canada)
4-Promotions
5-Administrative Appointments
6-New Appointments
7-Leaves for Special Appointments (NOT Sabbaticals)
8-Resignations
9-Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class): $12.00, regular AEA members and institutions
$ 6.00, junior members of AEA

All other countries (air mail): $18.00, regular AEA members and institutions
$12.00, junior members of AEA

Please begin my issues with:
☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name______________________________
First Middle Last

Address______________________________

____
City State/Country Zip/PostalCode

Check one:

☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

PREVIEW 1980

New Books from M. E. Sharpe, Inc.

Economic Democracy
The Challenge of the 1980s
By Martin Carnoy and Derek Shearer
A discussion of and an argument for alternatives to the present structure of the American economy—alternatives that would change the control of capital and how it is used.
"... an eye-opening tour through a series of options that informed citizens should know more about." —*Robert L. Heilbroner, New School for Social Research*

"... an excellent agenda for the discussion of that perennial question. What is to be done?" —*Robert Lekachman, City University of New York*

450 pages Cloth $15.00 Paper $7.95

The Megacorp and Oligopoly
Micro Foundations of Macro Dynamics
By Alfred S. Eichner
"... an extraordinarily thoughtful and skillfully executed theoretical treatise on modern corporate society.... Economic theorists, students of public policy and economic historians will find the book of special interest..." —*Business History Review*

"...Without doubt this is a stimulating and well-written book which deserves a wide audience." —*Economic Journal*

378 pages NOW IN PAPER $8.95

Industry and Business in Japan
Edited with an introduction by Kazuo Sato
Although much has been written by Japanese economists about Japan's rapid and tremendous economic growth since World War II, relatively little concrete and scholarly analysis of Japanese industry and business has been available in English. This volume fills the gap by gathering a number of important papers written by Japanese scholars on such topics as Japan's industrial growth, dual structure, industry studies, business groupings, and industrial policy.
450 pages $30.00

The Advanced Capitalist System
A Revisionist View
By Lynn Turgeon
A iconoclastic survey of the contemporary capitalist economy — defense spending, government budgets, markets, international trade, the dollar, labor, inflation and unemployment. Punctuated with provocative questions from the author's students at Moscow State University, this book will wake up those who doze over the standard texts in Economics I.
192 pages Cloth $15.00 Paper $7.95

Unemployment and Inflation: Institutionalist and Structuralist Views
A Reader in Labor Economics
Edited with an introduction by Michael J. Piore
"... provides an important antidote and challenge to neoclassical labor market explanations of inflation and employment... The quality of the individual selections is excellent. It's refreshing to be able to escape from the make-believe world of neoclassicism." —*David M. Gordon, New School for Social Research*
256 pages Cloth $20.00 Paper $7.95

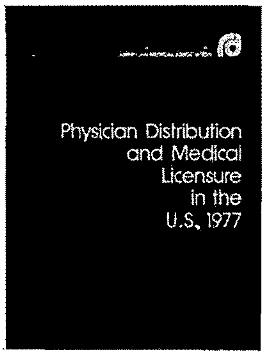
American Capitalism
The Concept of Countervailing Power
By John Kenneth Galbraith
With a new preface
"... London, Cambridge and Oxford together would, indeed, be hard put to point to any recent attempt at a general reformulation of basic economic doctrine carried out with such verve and skill as Mr. Galbraith's essay." —*The New York Times*

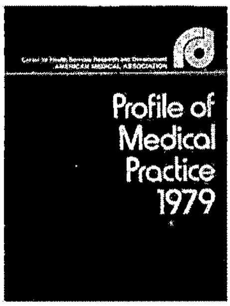
"(The book has) a profoundly satisfactory style, easy, lucid, witty, yet compact as befits the compactness of his mental processes." —*American Economic Review*

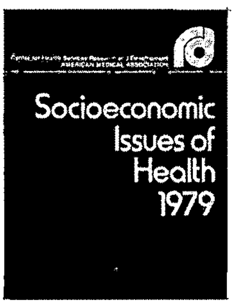
222 pages $17.50

For a complete catalog of books and journals, please write to:

M. E. Sharpe, Inc. 901 N. BROADWAY, WHITE PLAINS, N.Y. 10603







Just off the press!

3 NEW VOLUMES CONTAINING THE MOST CURRENT AND COMPREHENSIVE INFORMATION ON U.S. HEALTH CARE!

These new publications from the AMA Center for Health Services Research and Development are invaluable references for individuals and institutions in the health care field.

Physician Distribution and Medical Licensure in the U.S., 1977 provides statistics on the geographic and specialty distributions of physicians in the U.S., plus detailed tabulations on women and foreign medical graduate physicians. The monograph presents physician manpower by region, division, state, SMSA, and county; hospital data, population, and income are also included.

Profile of Medical Practice, 1979 contains data from the AMA's most recent Periodic Survey of Physicians. The volume profiles physicians' work patterns, fees, expenses, and income by specialty, census division, and other characteristics. Also featured are articles on medical economics, including papers by Frank A. Sloan and Joseph P. Newhouse.

Socioeconomic Issues of Health, 1979 features articles on the health care cost containment issue. Noted authors such as Alain Enthoven and Clark Havighurst examine the structure and possible cost impact of alternative health care delivery systems. Included are 56 tables of data on socioeconomic characteristics of the U.S. health care system.

Order your copies today!

Order Dept., American Medical Association AER
535 N. Dearborn St., Chicago, IL 60610

Please send me the following:
( ) **Physician Distribution and Medical Licensure in the U.S., 1977.** OP-054. $15 each.
( ) **Profile of Medical Practice, 1979.** OP-069. $6 each.
( ) **Socioeconomic Issues of Health, 1979.** OP-070. $6 each.
Enclosed is my check, payable to AMA, for $ ________.

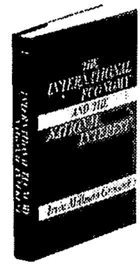
Please Print

Name ______________________________

Address ______________________________

City/State/Zip ______________________________

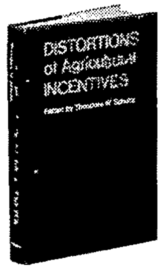
The International Economy and the National Interest



By Irvin Millman Grossack

In integrating theories of international trade and investment into a single model, Grossack both explains international investment in a clear and consistent manner and assesses its consequences for the welfare of individual nations. **$17.95**

Distortions of Agricultural Incentives



Edited by Theodore W. Schultz

"Highly recommended for . . . researchers in economic development and world agriculture, this work contains essays by leading economists and other authorities. . . . Many of the papers include firsthand observations and perceptions on problem situations, and original, hard-to-get data." —*Choice* **$12.95**

Available at bookstores or send $1.50 postage and handling for first book, 25¢ for each additional book, to order from publisher.

INDIANA UNIVERSITY PRESS Tenth and Morton Streets • Bloomington, Indiana 47405

PRAEGER

THE ECONOMICS OF MINERAL EXTRACTION
GERHARD ANDERS, Ontario Ministry of Natural Resources, W. PHILLIP GRAMM, S. CHARLES MAURICE, and CHARLES W. SMITHSON, Texas A & M University
ca. 350 pp. May 1980 LC 79-22949 ISBN 0-03-053171-3
ca. $24.95

MARKETING IN A SLOW-GROWTH ECONOMY
The Impact of Stagflation on Consumer Psychology
AVRAHAM SHAMA, Baruch College
184 pp. February 1980 LC 79-26312 ISBN 0-03-052151-3
ca. $19.95

THE POLITICAL AND SOCIAL ECONOMY OF COMMODITY CONTROL
CHRISTOPHER P. BROWN
ca. 400 pp. March 1980 LC 79-88568 ISBN 0-03-053351-1
$39.95

HEALTH ECONOMICS AND DEVELOPMENT
STUART WELLS, San Jose State University, and STEVEN KLEES, Universidade Federal, Brazil
ca. 180 pp. March 1980 LC 79-24821 ISBN 0-03-055796-8
ca. $18.95

GEOTHERMAL ENERGY AND REGIONAL DEVELOPMENT
The Case of Imperial County, California
edited by STAHRL W. EDMUNDS and ADAM Z. ROSE, The University of California, Riverside
398 pp. 1979 LC 79-19219 ISBN 0-03-053316-3 $26.95

MANAGING GROWTH IN THE 1980's
Toward a New Economics
ROBERT D. HAMRIN, Environmental Protection Agency
318 pp. February 1980 LC 79-19459 ISBN 0-03-054061-5
ca. $19.95
Paperback edition: ISBN 0-03-054056-9 ca. $8.95

PRAEGER
Praeger Publishers, CBS Educational and Professional Publishing, A Division of CBS, Inc.
521 Fifth Avenue, New York, New York 10017
(212) 599-8400

ANNOUNCING... NEW EDITIONS IN ECONOMICS

INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE, Second Edition

Frederic M. Scherer, Northwestern University

Extensively revised and rewritten to incorporate the most recent theoretical and empirical developments in the field. This edition retains all of the strengths of the original text, including the solid research foundation with excellent application of theory in real-world business contexts and a highly readable style, and now features...

- **an original theory of optimal product variety**
- **a new chapter on the dynamic theory of optimal limit pricing**
- **important new insights into the economics of product differentiation, economies of scale, and pricing behavior**
- **and, a new statistical appendix**

1980 632 pages Cloth

U.S. ECONOMIC HISTORY
Second Edition

Albert W. Niemi, Jr., University of Georgia

Topically organized within a chronological framework. This major revision focuses on the economic forces and issues underlying American economic development. Nine new chapters have been added to this edition, making the coverage even more comprehensive than before. Other chapters have been rewritten and updated to include the most current empirical research. The text is highlighted by the addition of 17 maps and 16 thought-provoking photographs.

1980 c. 512 pages Cloth

MODERN ECONOMICS: Principles and Policy,
Second Edition

Kelvin Lancaster & Ronald A. Dulaney

1979 803 pages Cloth

Also available in two volumes that present separate treatments of macroeconomic and microeconomic principles and policies.

RAND McNALLY COLLEGE PUBLISHING COMPANY
Box 7600 / Chicago, Illinois 60680

Give your students a sound introduction to economics...

Introductory Economics, Fourth Edition

Sanford G. Gordon, Associate Vice Chancellor for Policy and Analysis, State University of New York, Central Administration
George G. Dawson, Empire State College
1980/Casebound/480 pages

This Fourth Edition explains economic principles and concepts in an explicit and lively manner. Four review models — demand and supply, long-run equilibrium, pricing the factors of production, and national income determination — appear at appropriate intervals, forming a brief synopsis of the most important theoretical concepts in the book. *Study Guide* by Philip P. Robbins, State University College, Oneonta, N.Y. *Instructor's Guide with Test Items.*

Microeconomics: The Analysis of Choice

Daniel A. Graham, Duke University
1980/Casebound/496 pages

Designed for the intermediate course, this text develops microeconomic theory and clarifies the relationships between the various topics that comprise this body of theory. The analysis is positive (scientific) with the exception of chapter 11 (Welfare Economics) which includes a complete discussion of normative economics. The decision making process (choice theory) is the unifying theme for the text. Within this unifying context, consumption decisions, production decisions and investment decisions are seen as similar topics differentiated only by the nature of the choice being considered. The text assumes a non-calculus level.

Current Issues in the American Economy, 1980-1981 Edition

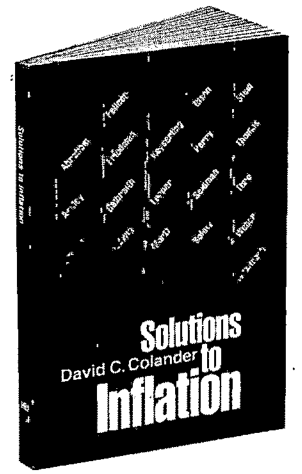
Edited by *Robert C. Puth,* University of New Hampshire
1980/Paperbound/320 pages

Completely updated to reflect the most current research and scholarship, this collection exposes the reader to a full gamut of opinions on issues selected for their timeliness, significance, and appeal to students in the introductory course. The book covers both macro- and microeconomic topics, making it a superb supplement in either one- or two-term principles courses. Each issue is developed through an introduction, discussion questions, a series of articles posing different views, and short annotated bibliographies.

125 Spring Street **Lexington, Massachusetts 02173**

Select a text from D. C. Heath!

Vital topics in economics...



MAP
A Market Anti-Inflation Plan

ABBA LERNER, Florida State University
DAVID C. COLANDER, University of Miami

An important original statement on a problem of universal concern, MAP suggests a solution to the problems of inflation/stagflation. The plan involves stabilizing average prices while allowing all individual prices and wages to adjust to changing tastes, technologies, and resources so that markets are cleared. MAP would achieve this through a mechanism that is not only fair to all parties and groups, but also avoids the well-known weaknesses of wage-price controls, monetary restraint, "jawboning," and fiscal restraint. The plan translates the social harm from the inflationary element in price and wage increases into a private cost that firms will try to avoid, using the market mechanism to evade administrative control or regulation of wages and prices.
Paperbound. 128 pages **Just Published**

Solutions to Inflation

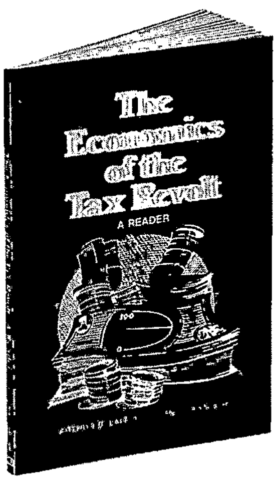
DAVID C. COLANDER, University of Miami

Is there a solution to the problems of inflation? This provocative collection suggests that perhaps there is, and provides 30 articles on current policy options—many of them written by their major proponents—for dealing with inflation. Among the topics discussed in the selections are indexing, monetary controls, wage-price guidelines and controls, tax-based incomes policies, and market-based anti-inflation plans. Some of the economists represented in the book are Friedman, Galbraith, Hayek, Solow, Okun, Ackley, Weber, Weintraub, and Lerner.
Paperbound. 224 pages.

HBJ

HARCOURT BRACE JOVANOVICH, INC.
New York • San Diego • Chicago • San Francisco • Atlanta

from today's leading economists and HBJ



The Economics of the Tax Revolt
A Reader

ARTHUR B. LAFFER,
University of Southern California
JAN P. SEYMOUR

This brief reader offers a series of 24 position papers and chronologies on five of the primary focal points of the national tax revolution: the economics of the Carter administration's energy plan; the Kemp-Roth 33% tax-cut measure; the Steiger-Hansen Bill that proposes a significant cut in capital gains taxes; California's Proposition 13, the Jarvis-Gann Initiative, limiting property taxes; and the "Laffer Curve," Professor Laffer's assertion that a large tax rate reduction would spur economic growth, increase tax revenues, and curb inflation.
Paperbound. 138 pages.

Coming in April

The Economics of Social Problems

RAY ROBINSON, University of Sussex
JULIAN LEGRAND, London School of Economics

Designed to show how economics can help us understand and overcome some of the pressing social problems in America today, *The Economics of Social Problems* presents an integrated, systematic framework for examining social issues ranging from health to the energy crisis. Following an introductory chapter that acquaints students with the concepts necessary for analyzing society's objectives and how a private market system may or may not realize these objectives there are nine chapters, each of which deals with a specific social problem.
Paperbound. 276 pages (probable)

HBJ

HARCOURT BRACE JOVANOVICH, INC.
New York • San Diego • Chicago • San Francisco • Atlanta

NORTH-HOLLAND ANNOUNCES

NORTH-HOLLAND PUBLISHING COMPANY
52 Vanderbilt Ave., New York, N.Y. 10017/ P.O. Box 211, 1000 AE Amsterdam, The Netherlands

NEW BOOKS • NEW BOOKS • NEW BOOKS • NEW

Policies for Employment, Prices, and Exchange Rates

Carnegie - Rochester Conference Series on Public Policy, Vol. 11

Edited by KARL BRUNNER, The University of Rochester, U.S.A. and ALLAN H. MELTZER, Carnegie-Mellon University, U.S.A.

1979. viii + 252 pages.
US $24.50/Dfl. 50.00 (paperback)
ISBN 0-444-85392-8

The eleventh Carnegie-Rochester Conference, held at Carnegie-Mellon University in November, 1978, features six papers. Two discuss theories of exchange rates and the evidence to support or reject particular theories. Two review the report of a committee of eight economists (the McCracken Committee) on the economic performance of the member countries of the Organization for Economic Cooperation and Development (OECD). The remaining two papers analyze some effects of specific employment and tax policies on the measured rates of unemployment and price changes.

Notes and Problems in Microeconomic Theory

Advanced Textbooks in Economics, Vol. 15

By PETER B. DIXON, La Trobe University, SAMUEL BOWLES, University of Massachusetts at Amherst, and DAVID KENDRICK, University of Texas at Austin, U.S.A.

1980. x + 318 pages.
US $36.50/Dfl. 75.00
ISBN 0-444-85325-1

This volume represents a unique approach to microeconomic theory, by setting up and solving main problems in mathematical programming, the theory of consumer demand, the theory of production, and welfare economics, and provides in-depth coverage of these topics. The book contains a set of material for supplementary work on, or self-teaching of, microeconomic theory. The material consists of annotated reading lists and sets of problems complete with extensively annotated solutions.

The Allocation of Road Capital in Two-Dimensional Space

A Continuous Approach

Studies in Regional Science and Urban Economics, Vol. 5

By TÖNU PUU, University of Umeå, Sweden

1979. xii + 218 pages.
US $41.50/Dfl. 85.00
ISBN 0-444-85324-3

This monograph describes applications of the *transportation capital* of a given total quantity should be allocated in two-dimensional space so that reasonable requirements for *economic efficiency* are met. The author uses a continuous transportation model with optimal paths determined by variational principles and with traffic flow regarded as a vector field. The mathematical analysis in this work is supplemented by extensive Appendices on vector and tensor analysis, complex variables, differential geometry, and the calculus of variations.

System-Wide Explorations in International Economics, Input-Output Analysis, and Marketing Research

LECTURES IN ECONOMICS: Theory, Institutions, Policy, Vol. 2

By HENRI THEIL, University of Chicago, U.S.A.

1980. x + 139 pages.
US $29.25/Dfl. 60.00 (paperback)
ISBN 0-444-85377-4

This monograph describes applications of the system-wide approach (or equation systems approach) to several areas, including consumer demand equation systems based on cross-country comparisons, import demand systems, sector demand systems based on an input-output model which includes price subsitution and demand systems for brands of the same product which include the effect of advertising. The measurement of the quality of consumption is also addressed and it is empirically implemented in a cross-country context.

0700 NH

"More Matter,

When Hamlet said this, he meant, "Let's get to the point." At Prentice-Hall, we've used this straight-forward approach to bring you and your students a great selection in economic texts. Our texts don't beat-around-the-bush. Instead, they provide students with hard-hitting facts and knowledge on the history of economics, modern economic theory, the economics of labor markets, price theory, and the future of the U.S. economy. To teach your students what really matters about economics, look these texts over today.

Economic Scenes: Theory in Today's World, 2/E
Stanley L. Brue and Donald R. Wentworth, both of Pacific Lutheran University

In a light, humorous style, this text explores selected micro and macro economic principles essential for an understanding of the functioning and malfunctioning of the American economy. Derives all economic concepts from scenes of human behavior and activity, and engages students in learning the theory and applications of economics through the use of fables, debates, and case studies.

1980 Paper 336 pp. est.

The Making of Economic Society, 6/E
Robert L. Heilbroner, The New School for Social Research

A much-acclaimed classic survey that traces the development of the market society and its problems from antiquity to the present. Chapters include the study of pre-market economy, the great depression, plus a current look at the problems of the 80's.

1980 Paper 336 pp. est.

Industrial Organization and Prices, 2/E
James V. Koch, Rhode Island College

A well-written undergraduate survey of Industrial Organization which presents both theory and application, as well as a great deal of empirical evidence. A review of relevant price theory opens the text, followed by an analysis of microeconomic theory.

1980 Cloth 528 pp. est.

Labor Economics: Theory, Evidence and Policy, 2/E
Belton M. Fleisher, Ohio State University, and Thomas J. Kniesner, University of North Carolina

This micro analysis of the economics of labor markets is designed to teach how the basic tools of microeconomic analysis and econometrics may be applied to labor market questions.

1980 Cloth 512 pp. est.

Economics of Poverty and Discrimination, 3/E
Bradley R. Schiller, American University

This text describes and analyzes recent policy in areas of manpower, welfare, and affirmative action. Added coverage of nonblack minorities, an expanded discussion of public and private pension and health systems, plus more are all included in this new edition.

1980 Paper 288 pp. est.

Price Theory and Applications, 2/E
J. Hirshleifer, University of California @ Los Angeles

A bestselling survey of modern price theory that introduces students to the use of verbal, geometric, and algebraic analytical approaches, with updated real-world examples and applied topics.

1980 Cloth 576 pp. est.

Money, Banking, and The United States Economy, 4/E
Harry D. Hutchinson, University of Delaware

Provides a broad range of monetary topics, including an analysis of the money supply, an in-depth study of the banking system, monetary theory, international finance, plus much more.

1980 Cloth 624 pp. est.

Prentice-Hall

For more information, or to reserve examination copies of these texts, please write to: Robert Jordan, Prentice-Hall, Inc., Dept. J-191, Englewood Cliffs, New Jersey 07632.

Less Art."

William Shakespeare

FOCUS ON 1980 ECONOMICS

A BASIC FRAMEWORK FOR ECONOMICS

Richard H. Leftwich, Oklahoma State University

This new text provides understandable, formal exposure to economic reasoning. It presents an internally consistent foundation of macroeconomic and microeconomic theory. Principles emphasized are those that are useful to students today, tomorrow, and in the future.

Student Workbook and Study Guide / Instructor's Manual

ECONOMICS OF SOCIAL ISSUES, Fourth Edition

Richard H. Leftwich and **Ansel M. Sharp,**
both of Oklahoma State University

Two new chapters have been added: "The Economics of Bigness" and "Government Expenditures and Government Revenues." The entire text has been thoroughly updated. It is the only issues book on the market that integrates issues and principles.

Student Guide and Workbook / Instructor's Manual

Examination copies for adoption consideration available on request: please indicate course title and text presently used.

BUSINESS PUBLICATIONS, INC.
Suite 1121 13773 North Central Expressway
Dallas, Texas 75243

The most successful new economics textbook of the decade

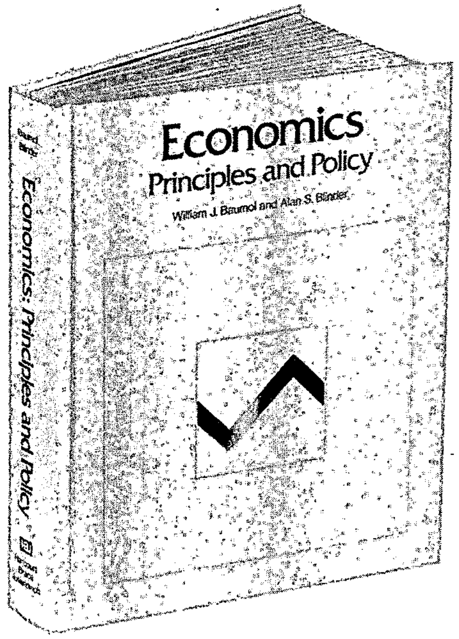
Economics
PRINCIPLES AND POLICY

WILLIAM J. BAUMOL,
New York University and Princeton University
ALAN S. BLINDER,
Princeton University

Just published last March and already in use at almost 200 colleges and universities, *Economics: Principles and Policy*, an introductory textbook, represents the best in scholarship and contemporaneity. *Economics* follows a traditional organization and begins by giving students a realistic understanding of macroeconomics. This section integrates price-level movement immediately and directly into the exposition. The unique development of aggregate demand and aggregate supply analysis allows an uncommonly clear explication of the employment-inflation trade-off, the stagflation dilemma, the Keynesian-monetarist debate, the burden of the debt, and other public-policy issues of great importance to today's economy.
862 pages.

"This is an excellent book,...which deftly integrates economic principles and applications without contrivance and a loss of integrity. While eschewing encyclopedic coverage, the book offers an excellent treatment of the crucially important topics so widely discussed by reasonably educated people in today's society."
PATRICK O'SULLIVAN, New York State University College, Old Westbury

"This looks like an excellent text, and it seems to take a fresh look at some of the real problems of our society. The chapter on the Keynesian-monetarist debate is especially crucial. Some of the examples, especially the regulation of taxicabs in Chicago, are very appropriate and should stimulate students' interest."
D. LARRY DAVIS, Indiana University, Gary



STUDY GUIDE
CRAIG SWAN, University of Minnesota
Provides, for each chapter of the text, learning objectives; a chapter review; basic and supplementary exercises; and self-tests.
Paperbound. 300 pages

Instructor's Manual, including chapter objectives, suggestions, questions and discussion questions for examinations.

Test Bank, 1,200 test items, available on request to instructors who adopt the textbook.

Transparency Masters reproduce 81 items from the textbook, including diagrams, charts, and curves; available on request to instructors who adopt the textbook.

HBJ

Harcourt Brace Jovanovich, Inc.
New York * San Diego * Chicago * San Francisco * Atlanta

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

WAUD IS HERE!

Economics

The First Brand-New Text for the 80's

"Recent shocks to the supply side of the economy, such as the rapid rise in imported oil prices, are making it painfully apparent that aggregate supply must be worked into macroeconomic analysis.

"This text shows the relationship between total demand and supply and illustrates why it is possible to have increases in the general price level and the unemployment rate at the same time. After standard Keynesian income analysis has been developed fully, the supply side of the economy is carefully integrated with it. This not only facilitates a comparison of Keynesian and Monetarist views, but also helps instructors deal candidly with cost-push inflation and the effects of stagflation on monetary and fiscal policy.

"Students expect, and should receive, a realistic analysis of such problems in an introductory course. We owe it to them." **—Roger N. Waud**

January 1980; 848 pages. Supplements: Instructor's Manual with Text Bank, Study Guide, Transparency Masters, and a computer-based examination system—ACCESS.

Two-volume paperback edition also available: *Macroeconomics,* 512 pages; and *Microeconomics,* 448 pages.

Isn't it about time you examined ECONOMICS for yourself—if you haven't already? Write today for an examination copy to Harper & Row, 10 East 53d Street, New York, N.Y. 10022. Please indicate course title, enrollment, and present text.

Harper & Row

New York • San Francisco

ECONOMICS

Michael E. Bradley, University of Maryland, Baltimore

At last, a principles text that challenges today's students...

- Complete four-chapter unit on unemployment and inflation, fully analyzing both supply-side and demand-side problems
- Detailed discussion of current issues and events: the economics of energy (full chapter), the drive for a "balanced budget" amendment, the 1979 Tokyo Round trade agreements, and the Three-Mile Island "incident"
- Concise, logical introduction to primary tools of microeconomic analysis: indifference curves, benefit-cost, input-output analysis, and isocosts/isoquants
- Full examination of latest macroeconomic theory: rational expectations, aggregate supply, the vertical Phillips curve, and the natural rate of unemployment
- Emphasis on the application of economic tools through both detailed examples and numerical problems

Available Now, 778 pages, illustrated, hardbound

Also available in two volumes:

Microeconomics and Macroeconomics

Available Now, 394 and 434 pages respectively, illustrated, softbound

Instructor's Manual with Test Items, Study Guide, and Transparency Masters are available

The revision you've been waiting for...

Urban Economics Second Edition

Edwin S. Mills, Princeton University

Available Spring 1980, 320 pages, illustrated, hardbound

For further information write
Jennifer Toms, Department SA
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman and Company

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

New
For International Economics

UNDERSTANDING INTERNATIONAL ECONOMICS: THEORY AND PRACTICE,

By J. David Richardson. Cloth/c. 528 pages/$16.95/with Instructor's Manual/now available for examination.

This new text offers an integrated, topical introduction to international economic analysis, policy, and interpretation — both "real" and financial — that requires a background *only* in introductory micro-economics and macroeconomics.

To help students understand the incentives underlying the strategies and actions in international exchange, Professor Richardson emphasizes the consequences of international economic transactions both for income distribution and for nations as a whole. In addition to standard coverage, he stresses material that is relevant to modern policy and private decision-making, including: *why* exchange rates and the balance of payments matter; external-currency (Euro-) banking; multinational corporations; and the new international economic order.

COLLEGE DIVISION
LITTLE, BROWN AND COMPANY
34 Beacon Street, Boston, Massachusetts 02106

PUBLICATIONS FROM THE UNITED NATIONS

STATISTICAL YEARBOOK 1978
Annual compilation providing comprehensive and accurate tables which group global aggregates and view world economic aspects in detail. Statistics relate to population, national income, agricultural and industrial production, energy, external trade and transport, communications, finance, health, housing, science and technology, education and culture.
Sales No. E/F.79.XVII.1 Clothbound $50.00

HANDBOOK OF INTERNATIONAL TRADE AND DEVELOPMENT STATISTICS 1979
Sales No. E/F.79.II.D.2 $44.00

BIBLIOGRAPHY ON TRANSNATIONAL CORPORATIONS. Centre on Transnational Corporations.
Sales No. E/F.78.II.A.4 $24.00

SCIENCE AND TECHNOLOGY FOR DEVELOPMENT. Terminology Bulletin No. 315.
Sales No. A/C/E/F/R/S.79.I.15 $26.00

UNITED NATIONS PUBLICATIONS
Room A-3315
New York, N.Y. 10017

UNITED NATIONS PUBLICATIONS
Palais des Nations
1211 Geneva 10, Switzerland

AVAILABLE NOW . . . FROM AEI

THE AEI ECONOMIST

Recent issues of this provocative monthly report have discussed such topics as the economic outlook for Great Britain under the Thatcher government, Iran's oil shutdown, the Carter administration's energy program, and views of the federal budget written by Herbert Stein and Gardner Ackley, both former chairmen of the President's Council of Economic Advisers. Contributors have also included Dr. Arthur Burns, former chairman of the Federal Reserve Board, and Professor Arthur Laffer of the University of Southern California.

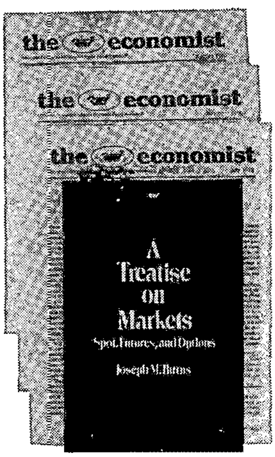
Edited and in large part written by Dr. Stein, senior fellow of the American Enterprise Institute for Public Policy Research and A. Willis Robertson, Professor of Economics at the University of Virginia, the **AEI Economist** is published monthly.

A TREATISE ON MARKETS

By Joseph M. Burns

The development of markets is perhaps the most important aspect of our free enterprise system. Yet little economic analysis of markets has been carried out. Burns' study deals with some fundamental principles of market development and breaks new ground by extending the work of Alfred Marshall on spot markets and that of Keynes and Hicks on futures markets, and by developing a theory of option markets.

145 pages / $4.75



☐ Enclosed is $1.00 for the next issue of the **AEI Economist**

☐ Enclosed is $10 for a one-year subscription (12 issues) to the **AEI Economist** (☐ Foreign airmail $16)

☐ Please send me ________ copies of **A Treatise on Markets**

Name ______________________

Address ______________________

City ______________________

State ______________ Zip ________

Payment must accompany orders under $10.

American Enterprise Institute
for Public Policy Research
Department 175
1150 Seventeenth Street, N.W., Washington, D.C. 20036

Please mention The American Economic Review *When Writing to Advertisers*

St. Martin's Press

Coming in '80

Introductory Statistics for Business and Economics

RICHARD K. MILLER, Washington and Jefferson College

Incorporating a wealth of examples from the real world, *Introductory Statistics for Business and Economics* consistently presents statistical methods as vehicles for making business and economic policy decisions. Professor Miller explains the rationale underlying the statistical techniques the student will use and has worked out each mathematical proof step by step. The coverage of *Introductory Statistics for Business and Economics* is extensive enough to meet the needs of any instructor teaching the introductory course; the organization is flexible enough to allow instructors to focus the course as they choose. The basic statistical concepts—descriptive statistics, probability, and statistical inference—are treated in the first eleven chapters. The remaining nine chapters offer a variety of topics from which to choose and may be covered in any desired sequence. The level of analysis requires only high-school algebra as a prerequisite. In addition to the numerous worked-out examples throughout the text, Professor Miller provides approximately 500 end-of-chapter problems. An appendix includes necessary statistical tables and answers to selected problems.

An Instructor's Manual provides answers to all problems, with complete solutions; additional problems; suggestions on use of the book; and chapter objectives.

A Study Guide includes chapter overviews and objectives; definitions of key words; new symbols and formulas; fill-in, self-testing questions; fully worked out examples with explanations; and problems with answers given.

February **500 pp.** (approx.); **graphs, charts, tables** **cloth $15.95**

Price and publication date of forthcoming books are tentative.
To request a complimentary examination copy, please write (specifying course, present text, and approximate enrollment) to:
St. Martin's Press P.O. Box 5352 New York, N.Y. 10017

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

Now available
Microeconomic Issues Today, Alternative Approaches

Coming in February '80
Macroeconomic Issues Today, Alternative Approaches

ROBERT B. CARSON, State University College, Oneonta, New York

Modeled on the author's highly successful *Economic Issues Today* (St. Martin's, 1978, $5.95), these two volumes are ideally suited for supplementary use in two semesters or quarters of the principles of economics course or in courses on contemporary economic issues. Incorporating recent developments and offering expanded coverage of topics and more data in the form of tables and charts, both books retain the features that made *Economic Issues Today* so teachable and so engaging for students:

- vital issues in current public-policy debate, clearly introduced and then argued from three different perspectives—the conservative, liberal, and radical
- lively, first-person advocacy of each position
- clear, simple language with only the necessary minimum of technical terminology
- Instructor's Manuals with discussion and short-answer questions and suggestions for using the books

Microeconomic Issues Today	**192 pp.**	**paper $4.95**
Macroeconomic Issues Today	**192 pp.** (approx.)	**paper $4.95**

Economic Issues Today will remain available for instructors preferring a concise edition combining microeconomic and macroeconomic issues in a single volume.

The Economics of Crime

DARYL A. HELLMAN, Northeastern University

In *The Economics of Crime* Professor Hellman presents a consistent framework for analyzing and developing public policy for dealing with crime. Her intention is to provide the student with a new perspective on crime and a general analytical approach which may be applied to new situations and new data. The first five chapters provide a theoretical framework; the succeeding five chapters analyze various categories of crime within that framework—crimes against property, crimes against persons, victimless crimes, the heroin trade, and organized crime. *The Economics of Crime* may be used in a variety of ways: as a supplement to an introductory course in microeconomics, as a foundation on which to build an advanced course in the economics of crime, or as one of the texts in an economic problems or issues course.

February 1980 **225 pp.** (approx.) **paper $5.95**

FROM

macmillan

New in 1980...

INDUSTRIAL ORGANIZATION AND PUBLIC POLICY

DOUGLAS F. GREER, San Jose State University

608 pages approx. ISBN 0-02-34702-8

The teaching of theory together with policy can be highly illuminating to the introductory student. This innovative industrial organization text is structured to provide that option.

It is also distinctly non-turgid and neither too long nor too encyclopedic.

It covers a substantial scope, including discussions of product differentiation, disclosure labelling, grade rating, regulation of product safety and quality, wage-price controls, and corrective advertising.

PUBLIC FINANCE AND THE PRICE SYSTEM

EDGAR K. BROWNING, University of Virginia
& JACQUELINE M. BROWNING, Sweet Briar College

464 pages ISBN 0-02-315650-3 1979

This text teaches students to apply theory in analyzing tax expenditure policies; it conveys the relevance and usefulness of economic analysis.

MONEY, BANKING, AND THE ECONOMY

Fourth Edition

JOHN A. COCHRAN, Arizona State University

588 pages ISBN 0-02-323040-1 1979

Completely up-to-date, this *Fourth Edition* offers balanced theoretical-institutional coverage of the economic issues and changes of the late 1970s—as well as an historical perspective.

Macmillan Publishing Co., Inc.

866 Third Avenue, New York, NY 10022

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

Industrial Concentration and the Market System

Legal, Economic, Social and Political Perspectives

The place of big business in our economy is the focus of the 50 articles and statements included in this compendium of current debate, opinion, theory, and practical experience. Drawn from a number of forums arranged by the American Bar Association and The Conference Board, the collection includes among its contributors:

Walter Adams	**Yale Brozen**	**Ira M. Millstein**
M.A. Adelman	**Alfred D. Chandler, Jr.**	**Barbara A. Reeves**
Betty Bock	**Theodore R. Eck**	**Frederick M. Rowe**
Robert H. Bork	**Allen C. Holmes**	**Robert L. Werner**

as well as a number of other distinguished economists, corporate and antitrust lawyers, historians, and policy planners.

A comprehensive introduction to the issue of industrial concentration, the volume can serve as a useful survey for practicing lawyers and economists as well as an excellent text for courses in industrial organization, antitrust law, and corporate law.

Edited by **Eleanor M. Fox** and **James T. Halverson**, "Industrial Concentration and the Market System" is published by the Section of Antitrust Law of the American Bar Association.

Paper $12.50 x, 311 pages 1979 6 x 9

Order Form

Please send me ____ copies of **Industrial Concentration and the Market System** (5030024) at $12.50* each.

Total (include $1.00 for handling) $____

Allow 4 to 6 weeks for delivery.

ABA

*$9.00 for law student members.

Name ____________

Address ____________

City ____________

State ____________ Zip ________

Make check payable to American Bar Association and mail to **American Bar Association,** Dept. 503, 1155 East 60th St., Chicago, IL 60637.

HOUGHTON MIFFLIN UPDATE:

Gregory/Stuart

Comparative Economic Systems

Paul Gregory, University of Houston
Robert C. Stuart, Rutgers, The State University of New Jersey
About 448 pages / Instructor's Manual / Just published

Gregory and Stuart provide a thorough, up-to-date, objective introduction. Their framework for identifying characteristics of economic systems helps students to analyze the relation between the characteristics and the performance of the systems. The authors thoroughly compare and contrast capitalist and socialist systems and their variants. Instructors can vary the analytic level of the text by adding to their courses more difficult material contained in the appendixes, and by omitting certain topics that are typographically set off within the text.

Marshall/Swanson

The Monetary Process: Essentials of Money and Banking, Second Edition

Robert H. Marshall
University of Arizona
Rodney B. Swanson, Security-Pacific National Bank, Los Angeles
About 496 pages / Instructor's Manual / Just published

Marshall and Swanson offer a balanced consideration of theory and policy for both the Keynesian and Monetarist schools. For this Second Edition the authors have given added emphasis to the industrial organization approach to commercial banking and have expanded their coverage of monetary theory. A comparatively brief text, Marshall/Swanson is suitable in introductory courses for both business and economics students.

Fletcher

Economics and Social Problems

Max E. Fletcher
University of Idaho
494 pages / Instructor's Manual
1979

Fletcher motivates students by presenting a number of contemporary economic problems and explaining the issues first. Then he applies economic analysis to the extent that it is useful in solving the problems, covering only those economic principles necessary for student understanding.

Gray

International Trade, Investment, and Payments

H. Peter Gray, Douglass College, Rutgers, The State University of New Jersey
669 pages / 1979

An innovative core text for advanced undergraduate and master's level courses in international economics. Gray's pragmatic text extends the traditional theoretical model of international trade and interrelates trade, investment, and payments to provide more realistic analysis. Includes extensive examination of multinational corporations and unusually thorough treatment of policy.

Ostrosky/Koch

Introduction to Mathematical Economics

Anthony L. Ostrosky, Jr.
Illinois State University
James V. Koch
Rhode Island College
371 pages / Solutions Manual / 1979

A one-semester text that covers the calculus and elementary linear algebra with strong emphasis on applications. The authors explain mathematical concepts and tools thoroughly, using numerous examples, then present plentiful economic and business applications after each major topic. Many problems and exercises, with answers, throughout each chapter help develop student comprehension.

Sampson/Farris

Domestic Transportation: Practice, Theory, and Policy, Fourth Edition

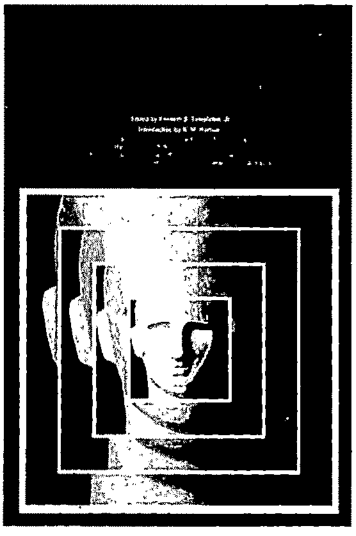
Roy J. Sampson
University of Oregon
Martin T. Farris
Arizona State University
518 pages / Instructor's Manual / 1979

For adoption consideration, request examination copies from your regional Houghton Mifflin office.

Houghton Mifflin

Dallas, TX 75234 / Geneva, IL 60134 / Hopewell, NJ 08525
Palo Alto, CA 94304 / Boston, MA 02107

LibertyPress
LibertyClassics



The Politicization of Society

Edited by Kenneth S. Templeton, Jr.

"The state has increasingly replaced the church in determining how we should behave," writes Oxford professor R. M. Hartwell in his introduction. "Politics is now religion." Fourteen scholars examine the central problem of modern society—the growth of the state—and its significance for the individual. They are Robert L. Carneiro, Felix Morley, Murray N. Rothbard, William Marina, Robert A. Nisbet, Jacques Ellul, Giovanni Sartori, Michael Oakeshott, Donald M. Dozer, Herbert Butterfield, John A. Lukacs, Jonathan R. T. Hughes, Butler D. Shaffer, and F. A. Hayek. Hardcover $10.00, Paperback $4.50.

We pay postage, but require prepayment, on orders from individuals. Please allow four to six weeks for delivery. To order this book, or for a copy of our catalog, write:
Liberty*Press*/Liberty*Classics*
7440 North Shadeland, Dept. 403
Indianapolis, Indiana 46250

Oxford University Press

International Economics
The Theory of Policy

GERALD M. MEIER, Stanford University. In the tradition of political economy, Professor Meier's text returns to the first principles of international economic policy, international payments policy, and international development policy, and focuses on the policy implications of international trade theory, international monetary theory, and international development strategies. The presentation avoids mathematics and complex diagrams. Optional appendices cover more specialized topics in the theory of international economic policy.

1980 350 pp.; 37 diagrams $16.95

World Development Report 1979

THE WORLD BANK. Foreword by **ROBERT S. McNAMARA.** The second in an annual series on world development problems, this report concentrates on such vital issues as energy, industrialization, employment, and urbanization. An extensive statistical appendix designed for easy reference shows selective economic and social data for more than one hundred countries.

1979 128 pp.; tables, charts cloth $9.95
paper $3.50

Money in International Exchange
The Convertible Currency System

RONALD I. McKINNON, Stanford University. Treating both the microeconomic and the monetary aspects of international exchange, this text describes how the present international system, based on purely national currencies, allows much of world trade to be monetized rather than bartered, and critically analyzes such financial practices as the selection of a currency of invoice, the extension of trade credit, and hedging against exchange-rate fluctuations.

1979 320 pp. cloth $12.50
paper $7.95

Prices are subject to change.

Classical and Neoclassical Theories of General Equilibrium
Historical Origins and Mathematical Structure

VIVIAN WALSH, The New School for Social Research, and **HARVEY GRAM,** Queens College and The Graduate School, City University of New York. Authors Walsh and Gram re-examine the primary sources of both classical (Petty, Cantillon, Quesnay, and Adam Smith) and neoclassical (Jevons, Menger, and Walras) views of the theory of general equilibrium, contrasting the simplest models that are capable of elucidating the fundamental differences between the two approaches, and providing the minimum structure needed for a clear look at certain substantial issues underlying the Cambridge controversy.

1980 448 pp.; 83 figs. $18.95

Philosophy and Economic Theory

Edited by **FRANK HAHN,** University of Cambridge, and **MARTIN HOLLIS,** University of East Anglia. The eleven essays in this collection explore three themes: the nature of economic explanation, the proper concept of rationality in economics, and questions of choice and value. Distinguished philosophers and economists consider the links between the abstractions of economics and such philosophical concerns as epistemology, ethics, and philosophy of science. *(Oxford Readings in Philosophy)*

1979 182 pp. paper $5.95

Industrial Economics
Theory and Evidence

DONALD A. HAY and **DEREK J. MORRIS,** both of the University of Oxford. A comprehensive review of recent theoretical and empirical works, this text emphasizes the need to relate the two. An innovative feature of the book is its integration of traditional analysis of the effect of market structure on performance with analysis of the behavior of large diversified firms that act to change market structures over time.

1979 664 pp.; 90 illus. cloth $33.00
paper $13.00

Oxford University Press 200 Madison Avenue New York, New York 10016

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

ECONOMICS... 1980 ... ECONOMI

ECONOMICS
PRIVATE AND PUBLIC CHOICE
Second Edition

JAMES D. GWARTNEY
Florida State University

RICHARD STROUP
Montana State University
with the assistance A. H. Studenmund, Occidental College

The turbulent 1970s have forced a new way of thinking about economics. This new edition of ECONOMICS: PRIVATE and PUBLIC CHOICE brings today's ideas about economics into your classroom. It takes an incisive look at what went wrong with the policies of the 1960s and shows how these policies led to the instability of the 1970s. Gwartney and Stroup analyze the importance of incentives in the economic system, explaining the contributions of the 'rubberband economists' your students read about in newspapers and magazines. They present an up-to-date examination of the critical issues affecting the energy market. And they show why 'stagflation' and declining productivity will continue to haunt the economic scene in the 1980s unless an economic policy course correction is made.
868 pages cloth

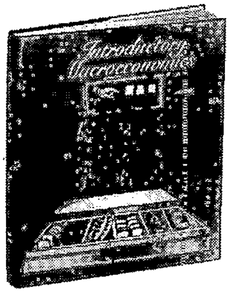
MACROECONOMICS: PRIVATE AND PUBLIC CHOICE, 2E
400 pages paper
MICROECONOMICS: PRIVATE AND PUBLIC CHOICE, 2E
400 pages paper

The Complete Package . . .

COURSEBOOK FOR ECONOMICS: PRIVATE AND PUBLIC CHOICE, 2E
A combined study guide and readings book.
300 pages paper

INSTRUCTOR'S MANUAL / TEST BANK / TRANSPARENCY MASTERS

CS... 1980 ... ECONOMICS... 198



INTRODUCTORY MACROECONOMICS

By MICHAEL VESETH
University of Puget Sound

Inflation and unemployment are analyzed in real-world settings using the aggregate demand—aggregate supply model of the economy.

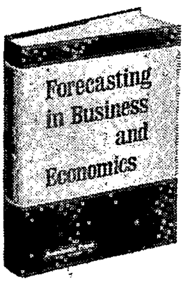
432 pages / paper   Instructor's Manual

ENERGY ECONOMICS AND POLICY

By JAMES M. GRIFFIN and HENRY B. STEELE
University of Houston.

The only textbook available on the impact of various government policies on the supply and demand for energy, particularly petroleum.

336 pages / cloth

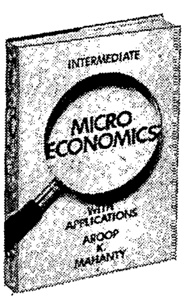


FORECASTING IN BUSINESS AND ECONOMICS

By C. W. J. GRANGER
University of California, San Diego

A junior-senior level treatment by the recognized authority on applying econometric theory to forecasting business cycles, author of the classic work FORECASTING ECONOMIC TIME SERIES.

250 pages / cloth



INTERMEDIATE MICROECONOMICS WITH APPLICATIONS

By AROOP K. MAHANTY
University of Northern Colorado

Emphasizes the immediate uses of microeconomic theory in every chapter through *Applications* sections that present mini-cases based on current business practice.

336 pages / cloth   Solutions Manual

Address requests for examination copies to: College Department, ACADEMIC PRESS, INC., 111 Fifth Avenue, New York, N.Y. 10003. Please state course title, enrollment, and present text.

ACADEMIC PRESS, INC.
A Subsidiary of Harcourt Brace Jovanovich, Publishers
NEW YORK LONDON TORONTO SYDNEY SAN FRANCISCO

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

New from WG&L...

International Financial Management

Edited by **Donald R. Lessard**
Associate Professor of Finance and International Management
Sloan School of Management, MIT

This new book brings together over forty of the most significant articles published in this vital area of financial management and written by an outstanding group of experts from universities and financial institutions. 1979.
6½" x 9½"   640 pp.   paper. $12.50

Financial Management of Commercial Banks

John M. Mason
The Wharton School, University of Pennsylvania

The first text to take a managerial approach to financial areas of bank administration, this new book focuses on the task of decision making within the banking firm and offers a unified framework for applying the most recent developments in corporate finance to a wide spectrum of bank activities. 1979.
464 pp. $16.50

Management Science in Banking

Edited by **Kalman J. Cohen**
Distinguished Bank Research Professor, Duke University
and **Stephen E. Gibson**
International Operations group, Chemical Bank

Presenting forty-three recent articles on the application of quantitative techniques to a wide range of bank administrative problems, this text clearly illustrates the potential of management science in dealing with the growing complexity of the commercial banking environment. 1978.   549 pp. $14.95

Modern Real Estate

Alvin L. Arnold, Editor, *Real Estate Review*
Charles H. Wurtzebach
University of Texas at Austin
and **Mike E. Miles**
University of North Carolina at Chapel Hill

Ready in February! This new text for the introductory course stresses the decision-making process in carrying out real estate transactions. As a result, it offers students an unusually practical and realistic view of the field today.
1980.

Send for your complimentary examination copies.
Please include course number and title, enrollment, semester taught, and present text.

WG&L ***College Textbook Division***
WARREN, GORHAM & LAMONT, INC.
210 South Street, Boston, Massachusetts 02111

Please mention The American Economic Review *When Writing to Advertisers*

The South-Western 1980 Line-up:

Keep an eye on us!

...keep an eye on us!

SW

SOUTH-WESTERN

Emerging Economics Publisher of the 80's

SOUTH-WESTERN PUBLISHING CO.

5101 Madison Road, Cincinnati, OH 45227
355 Conde Street, West Chicago, IL 60185
11310 Gemini Lane, Dallas, TX 75229
925 Spring Road, Pelham Manor, NY 10803
855 California Avenue, Palo Alto, CA 94304

St. Martin's Press

Public Utility Economics

Michael A. Crew and Paul R. Kleindorfer

Public Utility Economics provides a comprehensive analysis of the economic theory underlying pricing and investment problems of public utilities, from the perspectives of both welfare maximization and applied institutional design. The book first reexamines the welfare foundations of public utility economics, with special emphasis on consumers' surplus as a measure of social welfare. The authors then analyze peak-load pricing, including traditional theory; multi-period, multi-plant, interdependent demand; stochastic demand and dynamic analysis. Recent developments in reliability and time-varying demands are also treated. Professors Crew and Kleindorfer consider regulation versus public ownership in light of recent developments in agency theory, expense preference, and the traditional Averch-Johnson analysis. They apply the theoretical framework to the electricity supply industry and the problem of energy conservation. Important features of the book include a full exposition of the interactions of technology and pricing policy and a linking of dynamic and static theories of pricing. *October 280 pp. $19.95*

Economic Analysis of Government and Related Theories

Alan Peacock

An internationally recognized economist with extensive academic and governmental experience, Peacock has here revised and expanded his most important articles on the economics of public policy, previously scattered in journals, reviews, and *festschriften*. In new introductory chapters he summarizes his thinking on the political economy of government, the economic analysis of government policies, and the role of economists in government.

November 256 pp. $25.00

The Generalisation of the General Theory and other Essays

Joan Robinson

Robinson's *The Rate of Interest and other Essays*, first published in 1952, was a pioneering attempt to extend Keynes's theory of employment into long-run development analysis; the present volume is a revision and expansion of that long out-of-print work, taking into account the economic trends of the past thirty years. Joan Robinson has also included a new introductory essay reconciling Keynesian analysis with the revival of classical teaching introduced by Piero Sraffa. *September 176 pp. $18.95*

The Economics of the Performing Arts

D. C. Throsby and G. A. Withers

Recent years have seen an explosion of performing-arts activity throughout the English-speaking world. At the same time, prices for theater seats have skyrocketed and financial pressures have threatened to close major ballet and opera companies. The professional performing arts have become reliant on philanthropy and public funding, and government support has provoked extensive debate. Throsby and Withers examine the economic issues of this debate and define the position of the arts in contemporary society, developing and testing against the evidence of modern experience basic theories of the behavior of arts firms, consumers, and patrons in the United States, Canada, Britain, and Australia. *October 350 pp. $37.50*

Available at 20% professional discount.

St. Martin's Press P.O. Box 5352 New York, N.Y. 10017

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

Oldřich Kýn/Wolfram Schrettl (Editors)
On the Stability of Contemporary Economic Systems

Proceedings of the Third Reisensburg Symposium.
1979. 460 pp. and 1 folded plate, Paperb. DM 98,–
(Mathematical Studies in the Social and Behavioral Sciences 3)

"Selected Contents: Clower, Manove, Bajt on *Inflation*. H. Grossman, Azariadis on the *Labor Market*. Tintner, Kýn on *Cycles*. Mundlak on Factor Market *Distortion*. Koizumi on *Disequilibrium*. Th. Marschak on the *Productivity of Leisure*. K. Sato on *Technical Change*. Sontheimer on *Environmental Control*. Zarembka on the *Cambridge Position*. Round-Table Discussions."

Angewandte Statistik und Ökonometrie
Applied Statistics and Econometrics

Edited by Gerhard Tintner, Pierre Désiré Thionet and Heinrich Strecker

7. **Earl O. Heady/Thomas M. Reynolds/Donald O. Mitchell**
   An Econometric Simulation Model of the U.S. Farm Sector and its Policies and Food Exports
   1978. 61 pp., Paperb. DM 22,–
8. **C.W.J. Granger/Allan P. Andersen**
   An Introduction to Bilinear Time Series Models
   1978. 94 pp., Paperb. DM 28,–
9. **Pierre D. Thionet**
   Quelques problèmes concernant les sondages
   1978. 137 pp., Paperb. DM 34,–

11. **Gerhard Tintner/J.N.K. Rao/Heinrich Strecker**
    New Results in the Variate Difference Method
    1978. 102 pp., Paperb. DM 28,–
12. **V.A. Sposito/W.C. Smith/C.F. McCormick**
    Minimizing the Sum of Absolute Deviations
    1978. 60 pp., Paperb. DM 22,–
13. **Wolfgang Wertz · Statistical Density Estimation – A Survey**
    1978. 108 pp., Paperb. DM 24,–
14. **Peter Hackl**
    Testing the Constancy of Regression Models over Time
    1979. Ca. 132 pp., Paperb. ca. DM 32,–

Vandenhoeck & Ruprecht · Göttingen und Zürich

$1+2=?

DEMAND FUNCTIONS AND THE SLUTSKY MATRIX

S. N. Afriat

"While Slutsky's problem has been around for quite a while, this book represents the first systematic attack on the problem. It makes a significant original contribution to the area of demand theory."--*Lawrence J. Lau, Stanford University*

The utility idea has had a long history in economics, especially in the explanation of demand and in welfare economics. In a comprehensive survey and critique of the Slutsky theory and the pattern to which it belongs in the economic context, S. N. Afriat offers a resolution of questions central to its main idea, including sufficient conditions as well.

Princeton Studies in Mathematical Economics, 7

$13.50

Princeton University Press Princeton, New Jersey 08540

Solicitation of Nominations for

JAMES L. BARR MEMORIAL AWARD

The James L. Barr Award is made biannually for the best scholarly work in applied, public economics by a junior faculty member.

Purpose

To encourage young faculty in the Social Sciences to address applied questions in the public sector using the methods and theories of economics, broadly defined, consistent with the highest academic standards. These purposes reflect the too-short professional career of James L. Barr (1942-1977).

Eligibility

All University faculty members, except those who received their Ph.D. degrees prior to 1973.

Nominations and Award

Nominations should be addressed to the Awards Committee and, for the 1980 Barr Award, should be submitted prior to April 1, 1980. Nominations should consist of a brief statement on behalf of the nominee and copy of the article or scholarly work that is the basis of the nomination. The Award carries with it a $500 cash prize. The winner of the 1980 James L. Barr Memorial Award will be honored at a dinner in May of 1980.

Address Nominations to:

James L. Barr Memorial
Award Committee
c/o John P. Crecine
Carnegie-Mellon University
Pittsburgh, Pa. 15213

Awards Committee:

Richard M. Cyert
Otto A. Davis
John P. Crecine
Ted Bergstrom
John E. Jackson

The Joint Council on Economic Education

announces a new TUCE —

TEST OF UNDERSTANDING IN COLLEGE ECONOMICS

■ What is the TUCE? It is the only nationally normed test with which teachers can evaluate student knowledge before and after giving a typical introductory course in college economics. In addition, the TUCE provides a research instrument for controlled experiments or investigations that deal with techniques of teaching or cognitive learning of economics.

The new TUCE, which replaces the widely used original of 1968, is designed to reflect more recent thinking in economics. Accordingly, it puts more emphasis on the monetary and supply-side approaches than did the 1968 version.

■ What does the new TUCE consist of? It consists of three tests, one on microeconomics, another on macroeconomics, and a third hybrid (micro/macro) exam particularly useful for one-semester introductory courses. The tests stress the application of economic principles and concepts to realistic situations rather than the recognition or recall of facts. Each test comes in two forms—for pretesting and for post-testing—of thirty questions each.

■ Who developed the new TUCE? Rendigs Fels, Phillip Saunders, and Arthur L. Welsh constructed and tested the TUCE. An advisory committee consisting of G.L. Bach, Lee J. Cronbach, the late Robert A. Gordon, Allen C. Kelley, Paul A. Samuelson, and George J. Stigler provided supervision.

■ When will the new TUCE be available? The tests and answer sheets will be published in the fall of this year. Initially, specimen packages will contain these materials only; as soon as the interpretive manual is completed, it will be added.

Inquiries about the TUCE should be sent to:

Program Director
Joint Council on Economic Education
1212 Avenue of the Americas
New York, NY 10036

GENERAL EQUILIBRIUM, GROWTH AND TRADE

Essays in Honor of Lionel McKenzie

Edited by JERRY R. GREEN and JOSE ALEXANDRE SCHEINKMAN

A volume in the ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS Series

When economic theory was in its formative stages, Lionel McKenzie developed a novel and rigorous approach to it which has continued to be highly influential. His main areas of concentration remain the central thrusts of modern analytical economics. This collection of essays in honor of McKenzie extends the literature in those areas of mathematical economics and economic theory in which McKenzie made major contributions.

• In the area of economic theory, there are papers about the theory of demand in which McKenzie's pioneering contribution of 1952 is still a landmark; on Metzler matrices, in which McKenzie's early work was among the first to show their usefulness; on the stability of economic equilibrium; and on general equilibrium theory.

• In the area of economic growth, the continuous time approach for optimal control problems is explored in several papers by McKenzie's former colleagues and students.

• In the area of international trade, the questions of factor price equalization and of trade utility functions are explored in detail.

The level of each of the papers is at the forefront of research in their respective fields.

1979, 428 pp., $39.50 ISBN: 0-12-298750-0

ECONOMICS AND HUMAN WELFARE

Essays in Honor of Tibor Scitovsky

Edited by MICHAEL J. BOSKIN

A volume in the ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS Series

Tibor Scitovsky's scholarly work has spanned virtually all of economics and has influenced economists of several generations and many nations. From his earliest papers on welfare economics and the Pigou effect through numerous articles and books on international trade, economic development, and welfare economics, to his recent research on the very boundaries of economics and its sister social sciences, his work combines the highest professional standards with a rare originality. The essays in this volume, by some of his colleagues, students, and friends, live up to the high standards he himself has set. The contributors include some of the world's best known economists.

1979, 448 pp., $35.00 ISBN: 0-12-118850-7

MULTINOMIAL PROBIT

The Theory and Its Application to Demand Forecasting

By CARLOS DAGANZO

A volume in the ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS Series

CHAPTER HEADINGS: An Introduction to Disaggregate Demand Modeling in the Transportation Field. Maximum-Likelihood Estimation: Computational Aspects. Statistical Aspects of Multinomial Probit Model Calibration. Prediction: Mechanical Aspects. The Statistical Interpretation of Predictions. Appendices. References. Index.

Summaries are included at the end of each chapter.

1979, 236 pp., $24.50 ISBN: 0-12-201150-3

RESIDENTIAL CONSUMPTION, ECONOMIC OPPORTUNITY, AND RACE

By FRANKLIN D. WILSON

A Volume in the STUDIES IN POPULATION Series

Using a variety of sources and techniques, this interdisciplinary volume explores three related sets of issues in urban theory and policy. *Residential Consumption, Economic Opportunity, and Race*:

• Looks at families' housing consumption patterns, tracing the relationships between households' selected demographic and socioeconomic attributes and patterns of residential consumption.

• Examines trends in the suburbanization of metropolitan population and employment, analyzing the change that occurred during the 1960s and assessing the impact of this change on central city–suburban differentials in commuting patterns and wages.

• Estimates the effects of labor and housing market discrimination on blacks and the relationship between these two types of discrimination.

1979, 240 pp., $17.00 ISBN: 0-12-757980-X

MEASURING UNDEREMPLOYMENT

Demographic Indicators for the United States

By CLIFFORD C. CLOGG

A Volume in the STUDIES IN POPULATION Series

Series Editor: H. H. WINSBOROUGH

CHAPTER HEADINGS: Introduction. Underemployment in the United States, 1969–1973. Information in the Measures of Invisible Underemployment. The Effect of Changing Demographic Composition on the Crude Indicators of Underemployment. Latent Class Structure in the Distribution of Labor Market Rewards. The Dependence of Labor Force Status on Age, Time-period, and Cohort. Trending and Forecasting Underemployment: Log-Linear Models and Population Projection. Underemployment in the Life History of Synthetic Cohorts. Some New Directions for Mathematical Demography. Conclusion. Appendices. References. Index.

Conclusions appear at the end of each chapter.

1979, 288 pp., $21.50 ISBN: 0-12-176560-1

CLASS STRUCTURE AND INCOME DETERMINATION

By ERIK OLIN WRIGHT

A Volume in the INSTITUTE FOR RESEARCH ON POVERTY MONOGRAPH Series

Erik O. Wright, a sociologist in the Marxist tradition, has made a systematic effort to bridge the gap between contending theoretical perspectives on class and the growing body of quantitative studies of social and economic inequality. His basic theme is that class, defined not as an aggregation of individuals but as positions within social relations of production, plays a central role in structuring income inequality in advanced capitalism. In order to investigate this theme, Wright offers a new strategy for decoding and operationalizing class structures of contemporary capitalist societies. He pays particular attention to those locations in the class structure, such as managerial and supervisory positions, that do not fit neatly into the traditional class categories of Marxist theory; they belong neither to the working class, the capitalist class, nor the self-employed petty bourgeoisie. He argues that in order to understand income inequality it is necessary to examine the specific structural mechanisms through which income is determined within each of these different class positions.

Working with data from the Michigan Panel Survey of Income Dynamics, the Survey of Working Conditions (1969), and the Quality of Employment Survey (1973), Wright undertakes an intensive empirical analysis of class as a determinant of income, comparing its effects with those of occupational status, and analyzing its interactions with education, race, and sex. His results demonstrate conclusively that class has a systematic and pervasive impact on income inequality.

1979, 288 pp., $21.00 ISBN: 0-12-764950-6

FINANCING BLACK ECONOMIC DEVELOPMENT

By TIMOTHY BATES and WILLIAM BRADFORD

A Volume in the INSTITUTE FOR RESEARCH ON POVERTY: POVERTY POLICY ANALYSIS Series

The violence and civil disorders of the sixties spurred strong interest in programs that encouraged Black business development as a partial cure for the economic and social ills that beset Black ghetto populations. How successful have these programs been? What form has Black business development taken, and what has it accomplished in the Black community? The authors explore these questions and examine the effectiveness of government as an active financial intermediary. They measure Black business progress, identify problems, and analyze past and future policies for expanding Black-owned firms.

1979, 224 pp., $16.00 ISBN: 0-12-081650-4
Paper, $7.50 ISBN: 0-12-081652-0

Continuation Orders authorize us to ship and bill each volume in a series, or "Advances" type publication automatically, immediately upon publication. This order will remain in effect until cancelled. Specify the volume number with which your order is to begin.

Send payment with order and save postage and handling charge.

Prices are subject to change without notice.

ACADEMIC PRESS, INC.

A Subsidiary of Harcourt Brace Jovanovich, Publishers

111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

CAMBRIDGE

—an expanding economics publishing program
—a continuing commitment to advanced and specialist books
—a new determination to communicate the most recent ideas to all economists and their students

Profitability and Unemployment
Edmond Malinvaud

An examination of the rationale for the current argument in economic policy which states that an adequate level of profitability is a precondition for avoidance of future unemployment. Malinvaud's analysis is built around a simple model of the Harrod-Domar type, and extends the study of fixed price equilibria through a consideration of a dynamic process whose major components are capital accumulation and shifts in income distribution. A co-publication of Cambridge University Press and Les Editions de la Maison des Sciences de l'Homme. $12.95

Cambridge Economic Handbooks
Economic Theory and Exhaustible Resources
P. S. Dasgupta and G. M. Heal

The authors' authoritative presentation is concerned both with the descriptive modeling of how these resources will be used under different economic regimes and with a normative analysis of optimal use policies. An important feature is their integrated presentation of the theory of exhaustible resources and microeconomic theory in general. The book includes a number of empirical studies and extensive reference to institutional and policy matters. Hardcover $37.50 Paper $14.95

The Investment Decision of Firms
S. J. Nickell

A careful exposition of the relevant theory and a survey of the empirical research into investment behavior. Nickell makes an original contribution of synthesis and organization.
Hardcover $29.50 Paper $9.95

Cambridge Surveys of Economic Literature
Public Choice
Dennis Mueller

A review of the major contributions to the public choice literature over the past three decades. Using the basic tools of the economist, Mueller explores such questions as: Why do governments exist? What are the important properties of majority rule? What is the difference between outcomes under representative government and those that result from direct democracy?
Hardcover $24.95 Paper $6.95

Microfoundations
The Compatibility of Microeconomics and Macroeconomics
E. Roy Weintraub
Hardcover $19.95 Paper $6.95

Trade Amongst Growing Economies
Ian Steedman

Professor Steedman offers a simple analysis of international trade, emphasizing the means of production, profits, and growth rather than consumer demand patterns and the scarcity of natural resources. Considered in detail are such aspects of trade as the determinants of the pattern of trade, the welfare gain from trade, and the effects of such policies as tariffs, quotas, taxes, and multiple exchange rates. $21.50

Cambridge University Press
32 East 57th Street, New York, N.Y. 10022

New from Columbia

The Economics of Sex Differentials

CYNTHIA B. LLOYD AND BETH T. NIEMI. This book lucidly addresses a central paradox of American society: that while the labor force participation rate of women has increased dramatically, there has been little or no fundamental change in the division of labor in the home, or rewards in the market place. Offering a critical analysis of the economic literature on sex differentials as it relates to the last thirty years, the book provides the first thorough documentation for the radical changes that have occurred in all dimensions of female labor supply.

360 pages, $16.50

Women in the Labor Market

CYNTHIA B. LLOYD, EMILY S. ANDREWS, AND CURTIS L. GILROY, EDITORS. Representing such broad, vital areas of concern as household decision-making, labor supply, occupational choice, wage differentials, and equal employment opportunity policies, the essays in this volume begin the task of unravelling some of the facts behind the role of women in the economy. 424 pages, $25.00

The Failure of the Franklin National Bank

Challenge to the International Banking System

JOAN EDELMAN SPERO. In this timely account of the collapse of the Franklin National Bank, Joan Spero probes the tangle of mismanagement and deception that brought about the downfall of the nation's twentieth largest bank in October 1974. Working from original documents and interviews with key figures in the Franklin episode, Spero shows how the fragile web of bank interdependence across national boundaries almost swelled into a worldwide economic calamity. *A Council on Foreign Relations Book.* 240 pages, $14.95

Announcing:

Foundations of Political Analysis

An Introduction to the Theory of Collective Choice

ROBERT ABRAMS. This concise and detailed treatment of collective choice theory offers the first practical explanation of the subject to poliical scientists who are not specialists in this area but who need to apply the theory in their empirical or theoretical work. 368 pages, $16.50

To order send check or money order to **Dept. JN** at the address below, including $1.30 per order for postage and handling.

COLUMBIA UNIVERSITY PRESS
Address for orders:
136 South Broadway, Irvington, New York 10533

Wiley Books Mean Business, But Not As Usual...

PUBLIC AND PRIVATE INVESTMENTS:
Socioeconomic Analysis
Henry M. Steiner

Here's a book that describes the correct methods of judging capital investments. It's the only text that attempts to link project analysis with traditional microeconomic analysis. The first part of the book deals with decision making when the price system is assumed to be working perfectly. Part two discusses the meaning of resource allocation under the price system, particularly when the price system is not working. The third part examines the socioeconomic analysis in the development of nations.
(0 471 01625-X) 1979
approx. 450 pp. $24.95 (tent.)

MULTINATIONALS UNDER FIRE:
Lessons in the Management of Conflict
Thomas N. Gladwin and Ingo Walter

For the first time, a book that presents a comprehensive theoretical and practical analysis of multinational corporate conflict in a managerial framework. Issues ranging from human rights, terrorism, environmentalism and consumerism to labor relations, tax avoidance, and bribery/corruption in international business are discussed in detail. In each case, the issue is analyzed in cause-and-effect terms from the vantage-point of the multinational enterprise. Concrete cases anchor the concepts in real-world events. A new method of conflict diagnosis, planning and policymaking is also presented.
(0 471 01969-0) Feb. 1980
approx. 848 pp. $27.95 (tent.)

FORECASTING METHODS FOR MANAGEMENT, 3rd. Ed.
Steven C. Wheelwright and
Spyros Makridakis
(Wiley Series on Systems and Controls for Financial Management)

Although most managers are aware of the need for improved forecasting, few are familiar with the range of techniques that have been developed and the characteristics that must be known to select the most appropriate technique in a given situation. Here's a completely revised and up-dated edition that covers a wide range of forecasting methods and describes how they can be applied in practice, but does not dwell on the theoretical questions behind their development.
(0 471 05630-8) March 1980
approx. 300 pp. $24.95 (tent.)

UNPRICED VALUES:
Decisions Without Market Prices
J. A. Sinden and Albert C. Worrell

This book deals with the problem of placing comparative values on benefits and costs for which there are no market prices (e.g., scenic beauty, wildlife preservation, human health, outdoor recreation, etc.). The book, which demonstrates how decisions can be made in a rational manner, is the first comprehensive and systematic treatment of information on this subject. The methods that have been developed are organized into logical groups and they are offered in such a way that they can be useful to workers in fields other than those in which they were developed. There is an analysis of the relationship between valuation and decision making, and a discussion of just how the nature of the decision determines the kind of comparative values that are needed and the valuation costs that can be justified. Includes examples taken from the environmental and natural resource area to show how the methods can be applied.
(0 471 02742-1) 1979
511 pp. $25.95

MATHEMATICAL PROGRAMMING:
Structures and Algorithms
Jeremy F. Shapiro

One of the first books to incorporate and integrate most of the central constructs and main results of mathematical programming in one volume. It features many of the important algorithms for solving mathematical programming problems. This book also shows how to cut administration costs in welfare plans and how to speed processing of claims. It emphasizes Lagrangean duality theory and related concepts of convex analysis and explains how mathematics (including linear algebra, analysis, and group theory) is applied in the analysis and solution of optimization problems.
(0 471 77886-9) 1979
388 pp. $23.95

Available at your bookstore or write to John Storojev, Dept. 7563

WILEY-INTERSCIENCE
a division of John Wiley & Sons, Inc.
605 Third Avenue
New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario
Prices subject to change without notice.

092 0-7563

INCREASE YOUR ASSETS

NEW FOR 1980

ECONOMICS

Eleventh Edition
Paul A. Samuelson, Massachusetts Institute of Technology
900 pages (tent.), $18.95 (tent.)
Study Guide, $7.50 (tent.), Instructor's Manual, Test Bank, Transparency Masters.

A **major** revision of the classic principles text. The focus is on the 1980's, with extensive coverage of stagflation; up-to-date and objective coverage of the debate among proponents of Keynesianism, monetarism, and rational expectations; integration of micro- and macroeconomics throughout; and new material on such topics as the OPEC problem, the search for new international order, and the economics of time, marriage, crime and punishment, and discrimination (both sexual and racial).

MONEY AND BANKING, Second Edition

Dudley G. Luckett, Iowa State University
672 pages (tent.), $16.95
Instructor's Manual.

Excellent balanced treatment of both institutional and theoretical material, with regression analysis used to show the economic relationships between these two aspects of money and banking—that's what has made this text so popular. The new edition features a more extensive discussion of international banking; the use of portfolio theory to explain the role of financial intermediaries; and coverage of such current topics as rational expectations, the floating exchange rate, recent research on bank holding companies, and the changing concept of money.

Prices subject to change.

MONEY AND THE FINANCIAL SYSTEM: Theory, Institutions and Policy

Raymond E. Lombra, James B. Herendeen, both of Pennsylvania State University, and **Raymond G. Torto,** University of Massachusetts—Boston.
528 pages (tent.), $16.95 (tent.), Instructor's Manual.

With a lively and clear writing style, this new text provides superb, balanced coverage of financial institutions, the role of money and the financial system in determining the health of the economy, and the nature and complexity of the policy process. It takes an integrated look at the economy through a sectoral approach, and it features the latest developments, including the historic October 6, 1979 Federal Reserve announcement on policy actions designed to improve monetary control. Numerous learning aids provided.

COLLEGE DIVISION
McGRAW-HILL BOOK COMPANY
1221 Avenue of the Americas New York, N.Y. 10020

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

Announcing

the Formation of The Association for Public Policy Analysis and Management

THIS ASSOCIATION has been created for the purpose of encouraging excellence in research, teaching, and practice in the field of public policy analysis and management, including decisionmaking and policymaking theory, quantitative and other methods for analyzing public policies and their implementation and management, as well as the application of policy analytic theory and methods to substantive policy problems.

IT IS an individual membership organization, founded by public policy programs at the University of California at Berkeley, Carnegie-Mellon University, the University of Chicago, Columbia University, Duke University, Harvard University, the University of Michigan, the University of Minnesota, the State University of New York at Stony Brook, the University of Pennsylvania, Princeton University, The Rand Graduate Institute, Syracuse University, the University of Texas, and Vanderbilt University; public management programs at Boston University and Yale University; and the Brookings Institution, Mathematica Policy Research, Inc., and The Urban Institute.

THE ASSOCIATION plans to convene an annual Research Conference, to publish a journal of record, and to sponsor other activities designed to foster interaction among those trained in the several disciplines and professions who have a major interest in public policy analysis and management.

THIS ANNOUNCEMENT is intended also to constitute a call for proposals from those researchers wishing to be considered for invitation to present a paper at the 1980 Research Conference, to be held in Boston on October 24th and 25th. Those interested should send their proposals no later than March 1, 1980, to Professor Richard Zeckhauser, 1980 Research Conference Program Chairman. His address is John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138.

RESEARCHERS, PRACTITIONERS, teachers and students who are interested in joining the Association may do so by writing Professor Jared Hazelton, Treasurer of APPAM, Lyndon B. Johnson School of Public Affairs, University of Texas, Austin, TX 78712, and attaching a check for 1979-1980 dues ($20.00 for individuals and $10.00 for students). Members automatically receive the Association's Annual Research Volume consisting of selected papers presented at the annual Research Conference. Please note that the annual dues after 1979-1980 are expected to be perhaps as much as double the amount this year, in order to reflect the charge to be made for the journal of record, once it begins publication.

PLEASE NOTE that institutional memberships are also available. Those interested in exploring such affiliation should write Professor Joel L. Fleishman, President, APPAM, Institute of Policy Sciences and Public Affairs, Duke University, Durham, NC 27706.

Economics: An Introduction to Analysis and Policy, 10th/E
George Leland Bach, Stanford University
1980 cloth 800 pp. est.

A classic through the 70's now addresses the challenges of the 80's.

The extensive revision of Bach's successful economics text provides students with an even better core of basic economic theory and concepts for the 80's. Completely updated, this edition includes revisions of three important areas: macroeconomics offers a contemporary synthesis of Keynesian-Classical theory to explain and help analyze today's economic events; expansion of microeconomic theory, notably government regulation and deregulation of business; new coverage on the taxpayer's revolt and how large government should be.

Forty-three cases give students the opportunity to apply their knowledge to everyday issues found in jobs, newspapers, and politics. An improved format features larger graphs, easier reading, and new chapter summaries that make Bach's 10th edition more accessible and thought-provoking than ever.

Also available in paperback split

Macroeconomics: Analysis and Application, 2/E
1980 paper 496 pp. est.

Microeconomics: Analysis and Application, 2/E
1980 paper 480 pp. est.

Prentice-Hall

For more information, or to reserve an examination copy of this text, please write to: Robert Jordan, Prentice-Hall, Inc. Dept. J-192, Englewood Cliffs, N.J. 07632

STAFF PAPERS

Published by the International Monetary Fund in March, June, September, and December. Summaries of each paper in French and Spanish, as well as in English.

Vol. 26 December 1979 No. 4

The SDR as a Basket of Currencies .... J.J. POLAK

Fixed and Flexible Exchange Rates: A Renewal of the Debate
JACQUES ARTUS AND JOHN H. YOUNG

Inflation and International Reserves: A Time-Series Analysis .... MOHSIN S. KHAN

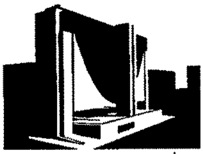
The Role of Foreign Direct Investment in the External Adjustment Process
DAVID J. GOLDSBROUGH

The Dynamics of Inflation: Forward Contracts and Money .... ARTURO BRILLEMBOURG

Monetary Policy in Selected Asian Countries
BIJAN B. AGHEVLI, MOHSIN S. KHAN, P.R. NARVEKAR, AND BROCK K. SHORT

Subscription: US$7.00 a volume; US$2.50 a single issue.
Special rate to university libraries, faculty, and students: US$3.00 a volume; US$1.00 a single issue.
Advice on payment in other currencies will be given upon receipt of order.

The Secretary, Attention: Publications Section, International Monetary Fund, Washington, D.C. 20431 U.S.A.



AVAILABLE MARCH 1980

Models of Monetary Economies

Proceedings and contributions
from participants of a December 1978 Conference
Sponsored by the Federal Reserve Bank of Minneapolis
John H. Kareken and Neil Wallace, Eds.

Contributors: David Cass, Masahiro Okuno, Itzhak Zilcha, Neil Wallace, James Tobin, Jose Alexandre Scheinkman, Martin Shubik, Robert E. Lucas, Jr., Leonid Hurwicz, Milton Harris, Frank Hahn, Truman Bewley, W. A. Brock, John Bryant, Karl Shell, Ross M. Starr, Robert M. Townsend

Softbound, 313 pp., $5.00
MN residents: add 4% sales tax;
check or money order must accompany order.
Foreign orders: Checks must be drawn
on U.S. banks, payable in U.S. funds.

Order from: Federal Reserve Bank of Minneapolis
Minneapolis, MN 55480
Attn: Office of Public Information

NEW! International Economics

Robert J. Carbaugh

Outstanding features:
- **Up-to-date analysis of issues and policies includes contending viewpoints**
- **Policy emphasis with close attention to theory and institutions**
- **Direct, readable style that is accessible to students**
- **Chapter outlines, summaries, and review questions focus on the important concepts**

Designed for Junior and Senior level courses this excellent new text combines theory, policy, and institutional material for an up-to-date coverage of modern international economics. The text provides numerous examples of actual economic problems along with statistical illustrations. Topics covered include modern trade theory, trade models, non-tariff trade barriers, commercial policies of the United States, OPEC, the monetary approach to balance of payments, and exchange rate systems.

Cloth; 448 pages; March, 1980

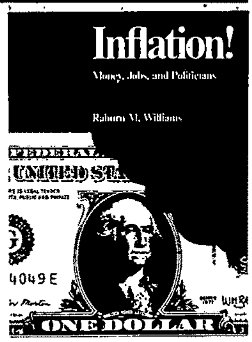
For further information, or to order examination copies, please write to Sara Black, Dept. 1980.

Winthrop Publishers, Inc.
17 Dunster St. • Cambridge, Mass. 02138

NEW TITLES FROM AHM

INFLATION!

Money, Jobs, and Politicians
Raburn M. Williams
University of Hawaii
about 152 pages / $4.95 (t) paper / January 1980



"I believe that this is the finest presentation of the economic history of macro policy and results for the last decade or so." Roger Miller,
University of Miami

Major revisions of two widely adopted texts:

CURRENT PERSPECTIVES IN BANKING

Operations, Management, and Regulation
Second Edition
560 pages / $13.95 (t) paper / January 1980

CURRENT ISSUES IN MONETARY THEORY AND POLICY

Second Edition
about 540 pages / $13.95 (t) paper / February 1980

Both edited by
Thomas M. Havrilesky
Duke University
John T. Boorman
International Monetary Fund

Please write for examination copies. For personal use, a 10% professional discount is allowed if payment is enclosed with order.

AHM Publishing Corporation
3110 North Arlington Heights Road
Arlington Heights,
Illinois 60004

In an age of economic confusion Dolan gives your students clarity.

BASIC ECONOMICS, Second Edition

EDWIN G. DOLAN,
George Mason University

In its first edition, Edwin Dolan's *Basic Economics* was the most successful introductory economics text of the last ten years. This new Second Edition further enhances the qualities that led to the book's great initial success—its readability, practicality, and concentration on today's economy and economics. In every chapter, the new edition reflects the very latest developments of fact, theory, and interpretation—monetary theory and policy, energy, balance of payments, rational expectations theory, and much more.

OUTSTANDING FEATURES:

- Emphasis on unemployment and inflation. (4 Chapters)
- Expanded coverage of market structure, anti-trust, and regulation.
- Appendices for difficult topics
- Real-world examples
- Case studies
- 3-way vocabulary development
- Tells how to read (and draw) graphs

DON'T FORGET THE PACKAGE:

- Study Guide (by Edwin Dolan)
- Instructor's Manual (by Edwin Dolan)
- Test Banks A and B
- Transparency Acetates
- Semiannual Newsletter

Now that the experiences of the 70s have shattered previous certainties, you need a text written for the 80s—Dolan's *Second Edition.*
January 1980, 736 pages,
ISBN 0-03-050356-6

ALSO AVAILABLE:

- **BASIC MACROECONOMICS: Understanding National Income, Inflation, and Unemployment, 2/e**
  February 1980, 400 pages, paper,
  ISBN 0-03-051276-X
- **BASIC MICROECONOMICS: Understanding Prices and Markets, 2/e**
  February 1980, 400 pages, paper,
  ISBN 0-03-051281-6

MACROECONOMICS

JACK VERNON, University of Florida

A focus on explaining the problems of unemployment and inflation and the role of government in dealing with them highlights this comprehensive and modern treatment of macroeconomics. The text incorporates the full employment constraint in each model as it is introduced, stressing throughout aggregate demand for output relative to full employment output as the key macroeconomic relationship. It features a careful, complete, and balanced account of the neo-Keynesian-monetarist debate, many applications and problems, and strong chapter introductions and summaries.
Available December 1979, 464 pages,
ISBN 0-03-042336-8

Dryden gives you 6 ways to fight economic uncertainty.

For complimentary copies, please send course title and approximate enrollment to:
James R. Ryder,
The Dryden Press, a division of Holt, Rinehart and Winston, 383 Madison Avenue,
New York, New York 10017

INTERMEDIATE MICROECONOMICS AND ITS APPLICATION Second Edition

WALTER NICHOLSON, Amherst College
A concise, well-organized development of basic theoretical concepts illustrated with numerous real-world applications.
1979, 614 pages, ISBN 0-03-041481-4

THE PRICE SYSTEM AND RESOURCE ALLOCATION

Seventh Edition
RICHARD H. LEFTWICH,
Oklahoma State University
Over half a million students have used this text to learn the fundamental principles of microeconomics.
1979, 487 pages, ISBN 0-03-045421-2

MANAGERIAL ECONOMICS

Third Edition
JAMES L. PAPPAS,
University of Wisconsin—Madison
EUGENE F. BRIGHAM,
University of Florida
This immensely successful book thoroughly covers the relationship between economics and managerial decision-making. Outstanding end-of-chapter problems.
1979, 656 pages, ISBN 0-03-045126-4

INFLATION AND UNEMPLOYMENT Second Edition

SAMUEL A. MORLEY,
Vanderbilt University
This supplement for Principles or Intermediate level macroeconomics and money and banking courses focuses on major theories of inflation and stabilization.
1979, 194 pages, ISBN 0-03-041016-9

PUBLIC ECONOMICS

Politicians, Property Rights, and Exchange
ADAM GIFFORD, JR., and
GARY J. SANTONI,
California State University—Northridge
How taxation and expenditure decisions are made in a democratic system and how they affect the allocation of resources. Extensive coverage of public choice.
1979, 273 pages, ISBN 0-03-021646-X

EXECUTIVE SEARCH

SPECIALISTS IN:

ECONOMICS
ECONOMETRICS
CORPORATE PLANNING
STRATEGIC PLANNING
MARKETING
OPERATIONS RESEARCH

STONEHILL IS AN EXECUTIVE SEARCH FIRM STAFFED WITH A TEAM OF PROFESSIONALS, EXPERIENCED IN SEARCH AND RECRUITMENT TECHNIQUES.

WE TAKE PRIDE IN OUR NATIONWIDE REPUTATION FOR HIGH STANDARDS OF PROFESSIONALISM IN EXECUTIVE RECRUITING.

OUR CLIENTS ARE "FORTUNE 500" COMPANIES, MAJOR CORPORATIONS, FINANCIAL INSTITUTIONS, AND CONSULTING FIRMS.

CONTACT: MR. GARY J. STEVENS, PRESIDENT
MS. DEE DROST, V.P. EXECUTIVE SEARCH

STONEHILL
MANAGEMENT CONSULTANTS, INC.
521 FIFTH AVENUE
NEW YORK, NEW YORK 10017

212 687-8150 **SUITE 1760**

LECTURES ON PUBLIC ECONOMICS
Anthony B. Atkinson, University of London, and **Joseph E. Stiglitz,** Princeton University
640 pages (tent.), $19.00 (tent.)

THE LABOR SECTOR, Third Edition
Neil W. Chamberlain, Columbia University, **Donald E. Cullen,** Cornell University, and **David Lewin,** Columbia University
669 pages, $13.95
Instructor's Manual

MACROECONOMICS, Sixth Edition
Thomas F. Dernburg, American University, and **Duncan M. McDougall,** Queens University, Canada
576 pages (tent.), $16.95 (tent.)

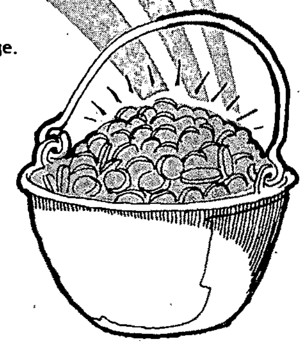
MICROECONOMIC THEORY: A Mathematical Approach, Third Edition
James M. Henderson, University of Minnesota, and **Richard E. Quandt,** Princeton University
492 pages (tent.), $18.95 (tent.),
Instructor's Manual

BASIC ECONOMIC PRINCIPLES, Fourth Edition
Robert A. Lynn, Kansas State University
544 pages (tent.), 15.95 (tent.),
Instructor's Manual

ECONOMIC ISSUES IN PUBLIC POLICIES
Richard B. McKenzie, Clemson University
320 pages (tent.), $8.00 (tent.)

PUBLIC FINANCE IN THEORY AND PRACTICE, Third Edition
Richard A. Musgrave, Harvard University, and **Peggy Musgrave,** University of California at Santa Cruz
800 pages (tent.), $19.00 (tent.),
Instructor's Manual

Prices subject to change.



A BONANZA OF NEW TEXTS FOR YOUR 1980 COURSES

COLLEGE DIVISION
McGRAW-HILL BOOK COMPANY
1221 Avenue of the Americas
New York, N.Y. 10020

STRIKE GOLD in the '80s

with Addison-Wesley's selection of outstanding economics texts. Like real gold, these books are a solid, valuable, high quality investment for you and your students.

ECONOMICS, THIRD EDITION

by Lewis C. Solmon

Solid coverage and balanced presentation

Systematic pedagogy

Successful for students and instructors

NEW

The New International Economics
by Ján S. Hogendorn, Colby College, and Wilson B. Brown, Northern Illinois University
(02824) 1980

Managerial Economics
by Pearson Graham, TRW, Inc. and Cleveland State University, and Diran Bodenhorn, Cleveland State University
(02470) 1980

Macroeconomics
by Robert D. Cherry, City University of New York, Brooklyn College
(00911) 1980

The Social System of the Planet Earth
by Kenneth E. Boulding, University of Colorado, Elise Boulding, Dartmouth College, and Guy M. Burgess, Massachusetts Institute of Technology
(00207) 1980

Economics: A Tool for Understanding Society
by Tom Riddell, Steve Stamos, and Jean Shackelford, all of Bucknell University
(06352) 1979

Economics: An Introduction to the World Around You, Second Edition
by Dennis J. Weidenaar and Emanuel T. Weiler, both of Purdue University
(08517) 1979

The Perspectives on Economics Series
Michael L. Wachter and Susan M. Wachter, University of Pennsylvania, Editors

New for 1980

International Trade
by Stephen P. Magee, University of Texas at Austin
(08365)

American Unionism: An Historical and Analytical Survey
by George H. Hildebrand, Cornell University
(08368) 1979

Money and the Economy: A Monetarist View
by William Poole, Brown University
(08364) 1978

The Economics of Medical Care
by Joseph Newhouse, The Rand Corp.
(08369) 1978

Income Distribution and Redistribution
by Paul Taubman, University of Pennsylvania
(08362) 1978

Development, the International Economic Order, and Commodity Agreements
by Jere R. Behrman, University of Pennsylvania
(08367) 1978

AntiTrust Policies and Issues
by Roger Sherman, University of Virginia
(08363) 1978

For complimentary copy consideration, write to Abigail Winslow, Addison-Wesley Publishing Company. Be sure to mention course title, enrollment, and text now in use.

Social Science & Humanities Division
ADDISON-WESLEY PUBLISHING COMPANY, INC.
Reading, Massachusetts 01867

afs

AUTOMATIC FORECASTING SYSTEMS
STATISTICAL CONSULTANTS

P.O. Box 563 / Hatboro, Pennsylvania 19040 / (215) 675-0652

"A Computer Program for the Analysis of Time-Series Models Using the Box-Jenkins Philosophy"

Author: Dr. David J. Pack
Version: September 15, 1978

BOX-JENKINS ANALYSES PERFORMED

I. Univariate times series model identification.
II. Univariate time series model estimation, diagnostic checking forecasting.
III. Single input transfer function model identification.
IV. Single input transfer function model estimation, diagnostic checking, forecasting.
V. Multiple input transfer function model identification (assuming independence of inputs).
VI. Multiple input transfer function model estimation, diagnostic checking, forecasting.
VII. Intervention model estimation, diagnostic checking, forecasting (Box-Tiao, **Journal of the American Statistical Association,** March, 1975).

A Program source on tape, including program subroutines, sample driving programs and sample data decks. The tape is a 9 track tape produced via an IBM 370/168 system with the following parameters: DSB=DATA, RECFM=FB, LRECL=80, BLKSIZE=12960, DENSITY=1600 bpi. **University rate $150.** ☐ **Non-University rate $300.** ☐

B Error notices, amendments, modifications and other advisories. Includes membership in Users Group with annual meetings. **Yearly costs only $25.** ☐

C Extensive telephone consultation to support your usage or to answer questions related to Box-Jenkins modelling including model selection. Includes all privileges of subscription B. **Yearly costs only $100.** ☐

D 150 page User's Manual. **$25 per copy.** ☐

PURPOSE AND DESCRIPTION OF AUXILIARY COMPUTER ROUTINES—Author: David Reilly

E Generate good starting values for Transfer Function Estimation. **$150.** ☐

F Express univariate Box-Jenkins model as weighted sum of the past (P i weights). **$250.** ☐

G Express Transfer Function and Noise model as ordinary regression equation. **$200.** ☐

H Compute variance of aggregated forecast (Univariate and Multivariate). **$200.** ☐

To: Automatic Forecasting Systems
P.O. Box 563
Hatboro, PA 19040

☐ Please send me more information about your programs.

☐ I have enclosed payment for the items checked above.

Name ______________________________

Organization ______________________________

Street Address ______________________________

City ____________________ State ____________ Zip Code __________

Begin making plans to attend the

Ninety-Third
Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

DENVER, COLORADO

September 5-7, 1980

The 1980 Professional Placement Service will be held December 28-30 at the Dallas Hilton, Dallas, Texas. Please note that there will not be a Professional Placement Service in Denver at the September meetings.

See the Notes section of the June *AER* for the American Economic Association's preliminary program.

The 1981 meeting will be held in Washington, DC, December 28-30.

IRWIN

All New Texts

PRINCIPLES OF ECONOMICS: MACRO
PRINCIPLES OF ECONOMICS: MICRO
4th Edition—Twin Paperbound Volumes
Willis L. Peterson, *University of Minnesota*

THE PUBLIC FINANCES: AN INTRODUCTORY TEXTBOOK, *5th Edition*
James M. Buchanan, *Virginia Polytechnic Institute and State University,* and **Marilyn R. Flowers**, *University of Oklahoma*

MONEY, BANKING, AND MONETARY POLICY
Douglas Fisher, *North Carolina State University*

PROBLEMS IN LABOR RELATIONS: TEXT AND CASES
William E. Fulmer, *University of Alabama*

MICROECONOMIC THEORY, *Fifth Edition*
John P. Gould, *University of Chicago*, and the late **C. E. Ferguson**

THE ECONOMICS OF DEVELOPMENT, *Third Edition*
Everett E. Hagen, Professor Emeritus, *Massachusetts Institute of Technology*

LABOR ECONOMICS: WAGES, EMPLOYMENT, AND TRADE UNIONISM, *Fourth Edition*
F. Ray Marshall, *U.S. Secretary of Labor,* **Allan G. King**, *University of Texas,* and **Vernon M. Briggs, Jr.**, *Cornell University*

READINGS IN LABOR ECONOMICS AND LABOR RELATIONS, *Fourth Edition* *(paperbound)*
edited by **Richard L. Rowan**, *University of Pennsylvania*

Examination copies for adoption consideration available on request.
Please indicate course title and text presently used.

RICHARD D. IRWIN, INC.
Homewood, Illinois 60430

9 JUL 1981

The American Economic Review

PAPERS AND PROCEEDINGS

OF THE

Ninety-Second Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

Atlanta, Georgia, December 28–30, 1979

MAY 1980

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

Officers

President
MOSES ABRAMOVITZ
Stanford University

President-Elect
WILLIAM J. BAUMOL
Princeton University and New York University

Vice Presidents
CARL F. CHRIST
Johns Hopkins University
H. GREGG LEWIS
Duke University

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review
GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature
MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee
ROBERT W. CLOWER
University of California-Los Angeles
MARINA v. N. WHITMAN
General Motors Corporation
HENRY J. AARON
The Brookings Institution
ZVI GRILICHES
Harvard University
MARTIN S. FELDSTEIN
National Bureau of Economic Research, Inc. and Harvard University
ROBERT E. LUCAS, JR.
University of Chicago

Ex Officio Members
TJALLING KOOPMANS
Yale University
ROBERT M. SOLOW
Massachusetts Institute of Technology

• Published at George Banta Co., Inc. Menasha, Wisconsin. The publication number is ISSN 0002-8282.

• *THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September, and December.

Effective April 1, 1980, Association membership dues for 1980, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$28.75 for regular members with rank of assistant professor or lower, or with annual income of $13,800 or less;

$34.50 for regular members with rank of associate professor, or with annual income of $13,800 to $23,000;

$40.25 for regular members with rank of full professor, or with annual income above $23,000;

$14.25 for junior members (registered students). Certification must be submitted yearly.

Subscriptions (libraries, institutions, or firms) are $43.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $4.75 to cover extra postage.

• Correspondence relating to the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

• Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212.

THE AMERICAN ECONOMIC REVIEW

VOL. 70 NO. 2 MAY 1980

PAPERS AND PROCEEDINGS

OF THE

Ninety-Second Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

Atlanta, Georgia

December 28–30, 1979

Program Arranged by MOSES ABRAMOVITZ

Papers and Proceedings Edited by GEORGE H. BORTS AND DANIEL F. SPULBER

Copyright American Economic Association, 1980

CONTENTS

PAPERS

PROCEEDINGS

THE purpose of the American Economic Association according to its charter, is the encouragement of economic research, the issue of publications on economic subjects, and the encouragement of perfect freedom of economic discussion. The Association as such takes no partisan attitude, nor does it commit its members to any position on practical economic questions. It is the organ of no party, sect, or institution. People of all shades of economic opinion are found among its members, and widely different issues are given a hearing in its annual meetings and through its publications. The Association, therefore, assumes no responsibility for the opinions expressed by those who participate in its meetings. Moreover, the papers presented are the personal opinions of the authors and do not commit the organizations or institutions with which they are associated.

Introduction–Program Chairman

The job of arranging a program for the Association's annual meetings is both easier and harder than it may seem to be. At the Atlanta meetings, the AEA appeared as the sponsor of some seventy sessions—a formidable number. Some thirty of those, however, were arranged jointly by the AEA and one or another of our allied associations, and in most though not all these joint sessions, the allied association took the initiative and was primarily responsible for arranging the meeting. I was able to arrange the forty-odd AEA sessions proper because, like every program chairman, I had help. I was able to appeal to people who were actively engaged in the investigation of one or another subfield to take responsibility for a session. Some of them were old friends and collaborators. In any case, I found a very ready response, and I think everyone should understand that the success of our meetings from year to year depends on the fact that many rather busy people are willing to accept the work and bother of planning a session, finding able contributors and discussants, and seeing to it that, on the appointed day, the participants are present, clutching their papers and comments and ready to read. Besides the invited papers that the *Proceedings* now publish, nine sessions were arranged on the initiative of members who proposed a subject and persuaded me it was interesting and that good papers could be obtained. Finally, nine other sessions were arranged to hear papers contributed by members in response to the "call for papers." All that generous effort brought us through the Atlanta meetings, and I should like to record my thanks for it here.

Because of the help one can obtain, arranging a program is relatively easy—so long as one is satisfied simply to aim for sessions that spread, in a more or less-balanced way, over the various fields of active work. It is harder to arrange a program which has some visible coherence—an apparent "theme." That is partly because such coherence conflicts with the need for enough balance and diversity to represent the varied types of research now going on. Thematic unity also faces the difficulty that potential contributors obstinately insist on preparing reports of the work they are actually doing and not of work which, in a program chairman's vision, they should have been doing. Forewarned about these facts of life, I did not press hard to impose a sharply etched private stamp on the sessions.

But neither did I maintain an utterly blank neutrality. If 1979 was the fiftieth anniversary of the onset of the Great Depression, by the same token it was the fiftieth anniversary of the watershed year which separates the era of unregulated market capitalism and minimal government from the mixed economy that followed. We have now clearly entered a period when the vitality, if not viability, of the new regime is being questioned and tested. Can we maintain the private sector's efficiency and dynamism while we use government to alter the market's verdicts and cushion its effects? That seemed to me to be a question worth exploring. The faithful reader of these *Proceedings* will, therefore, find sessions that bear on a variety of issues on which the answer to my central question will turn. Here is a short list, headed by Tibor Scitovsky's insightful lecture.

Can Capitalism Survive?—An Old Question in a New Setting (*The Richard T. Ely Lecture*)

The Government and Capital Formation

Regulatory Intervention in Historical Perspective

Self-Correcting and Self-Aggravating Forces in Urban Decline

A General View of Capital Formation and Economic Growth

The Current Retardation in Productivity Growth

The Political Economy of National Health Insurance

The Evolving World Dollar Standard

The Effects of Fiscal Policies on the

Distributions of Income and Wealth
Taxation, Labor Supply and Saving
The Recoil from Welfare Capitalism: Political and Sociological Perspectives

When Robert Solow last year accepted George Borts' invitation to write a prefatory page in explanation, and perhaps exculpation, of his program, it was the first time it had been done. He intoned, "Thus are precedents created." But what kind of precedent would it have been had no second introductory page ever appeared? Here it is.

MOSES ABRAMOVITZ

Editors' Introduction

This volume contains the *Papers and Proceedings* of the ninety-second annual meeting of the American Economic Association. The *Proceedings* consist of the record of the business activities of the Association in 1979: the annual membership meeting; the March and December meetings of the Executive Committee; and reports of the Association's officers and committees. As with the notes section in each issue of the *American Economic Review*, they are published to keep the members informed and encourage them to participate in the Association's affairs.

The *Papers* constitute the greater part of this volume. They comprise sixty-six papers and twenty-five discussions, and are roughly equivalent in space to two regular issues of the *American Economic Review*. They are, however, published under different procedures. About a year in advance, the Association's President-elect (in 1979 Moses Abramovitz, in 1980 William Baumol) acting as program chairman, decides on the topics for which sessions will be organized. This is done after consultation and comment, both volunteered and solicited from a wide range of individuals. The program chairman sets limits on the length of the papers at various sessions, and invites persons to organize these sessions. Each session organizer in turn invites several persons (usually two or three) to give papers on the theme of the session, and asks others to give comments on the papers. The program chairman decides at the time of organization which sessions are to be printed in this volume. Space limitations restrict the number of printed sessions. This year we are printing twenty-three sessions, although a total of seventy sessions were sponsored, either solely by the American Economic Association or jointly with other allied societies.

There is no standard practice with regard to the publication of comments and discussions, and each program chairman must decide how to allocate the publication space between invited papers and discussions. In the present volume we are publishing twenty-five comments. Not all of the comments received were of publishable quality, and it fell to the session chairmen as well as the editors to decide which to use.

The rules under which these papers are published are quite different from those governing the regular issues of the *Review*. Their length is strictly controlled. Except in unusual circumstances, they must be less than 4,000 (sometimes 3,000) words in length. Their content and range of subject matter reflect the wishes of the program chairman to investigate and expose the current state of economic research and thinking. In many cases, they are exploratory and discursive rather than definitive presentations of research findings. Because they may represent conjecture and expectation, they provide unusual insight into the state of mind of the authors. While we edit the papers to improve content and style, to satisfy space requirements and eliminate repetition, we do not subject the papers to a refereeing process, and publication of any paper received prior to the printing deadline that satisfies space requirements is virtually guaranteed.

We would, however, refuse to publish a paper if we concluded after reading it that it was utterly without merit; no paper has yet been rejected on these grounds. The Executive Committee has established another ground for rejection: if a paper cannot be cut to meet space requirements, we may ask the author to allow its consideration for publication in a regular issue of the *Review*, subject to the usual refereeing process, or the author may be asked to withdraw the paper and submit it elsewhere.

These practices serve a number of important purposes: the papers can be published without the long delays imposed by the refereeing process. They are short papers, covering a wide variety of subjects, and in most cases can be understood by nonspecialists. (Indeed they provide excellent text material for certain teaching purposes.) Authors receive a chance to report on

research to be undertaken or recently completed, discuss topical subjects in an informal way, and summarize longer forthcoming publications. Readers get a chance to browse among a large number of articles which are outside their major areas of interest, but which are not as specialized or as technical as those sometimes found in the regular journals. And while the papers are not refereed, they do provide an accurate picture of the state of thinking in many of the fields of economics.

GEORGE H. BORTS
DANIEL F. SPULBER

RICHARD T. ELY LECTURE

Can Capitalism Survive?—An Old Question in a New Setting

By TIBOR SCITOVSKY*

Can capitalism survive? That was the question raised a generation ago by Josef Schumpeter and answered in the negative. We have accumulated a lot more hindsight since then. Even so, you would hardly expect me, a senior-citizen economist from California, the retirement state, to be so bold and bright as to try to improve upon the most celebrated performance of America's most brilliant economist. Nor, for that matter, would I wish to do so even if I could—certainly not here, where I am addressing an audience of professional forecasters and would-be forecasters, and not at this time, when the sheer volume and variety of economic forecasts are depreciating the product. I propose instead merely to change the venue. Not to reconsider Schumpeter's argument and conclusions but to approach the question differently and see what additional light that throws upon it.

Capitalism, when you think of it, is not an attractive form of social organization; but it is, or at any rate has been, redeemed by two great merits: the impersonal nature of its constraints; and its unequalled flexibility —flexibility in exploiting opportunities, absorbing shocks, adapting to changed circumstances. I shall focus on this last feature and ask whether that flexibility still exists and still redeems.

To come back to the variety of forecasts for a moment, the remarkable thing about them is not the variety itself but the fact that those many different forecasts all have the same basis. We economists proudly consider our discipline the queen of the social sciences, on the ground that unlike other social scientists, we subscribe to the same paradigm and use the same concepts, approach, and tools of analysis. All the more surprising therefore is the contrast between the monolithic solidity of the base we start from and the great variety of predictions, diagnoses, and policy recommendations we end up with. That, I dare say, has many reasons; but there is one, which to my mind, is more important than the rest, and which is well exemplified in the field of international money.

In the good old days, when there still were believers along with the unbelievers in the functioning of the international financial system, balance-of-payments theorists used to be divided into elasticity optimists and elasticity pessimists. They shared the same theory of balance-of-payments adjustment, which said that a system of variable exchange rates would be self-equilibrating if the Marshall-Lerner conditions were satisfied, but disequilibrating if the demand and supply elasticities of imports and exports were too low to satisfy those conditions. The optimists and the pessimists differed merely in attributing different values to those elasticities, whose actual values no one knew nor had means of finding out.

Another example of a similar division of opinion comes from the theory of growth. There used to be a distinguished group of economists whom you might have dubbed the growth pessimists. Alvin Hansen and other believers in his stagnation thesis were in that group; and so were Sir Roy Harrod and Hansen's pupil, Evsey Domar, who in a classic case of tacit collusion propounded

*University of California at Santa Cruz and Stanford University. I am grateful to Moses Abramovitz, Nathan Rosenberg, Stanley Sheinbaum, and Lorie Tarshis for many useful suggestions and corrections; but the neck I am sticking out is entirely my own.

the Harrod-Domar growth model, so elegant in its simplicity and so disturbing in the instability of its knife-edge equilibrium. Fortunately, there was also a no-less-distinguished group of growth optimists, led by Robert Solow and Trevor Swann, who gave us the equally elegant but reassuringly stable Solow-Swann model of economic growth. Here, too, the difference between optimists and pessimists lay in the differing assumptions they made about elasticities. Indeed, we could again call the pessimists elasticity pessimists, since their dire conclusions hinged on their assuming fixed coefficients and zero elasticities.

That similarity between the two examples is not accidental, of course. On the contrary, they are both variations on the same simple theme, which is that capitalism works when it's flexible, but self-destructs when it's not. The great merit of capitalism and the secret of its past successes was its great flexibility, imparted to it by the automatic forces of the market.

What then is that flexibility and how does it come about? In general, there are two kinds of economic flexibility. The first is the flexibility of the economy's individual members: the flexibility of consumers in adapting their expenditure patterns, of producers in modifying methods of production and utilizing the most available outputs, and of the owners of productive services in providing the services most in demand. The second and more drastic kind is the ability of the system as a whole to adjust its reliance on its individual members: increasing its reliance on those best adapted or able to adapt to changed circumstances and jettisoning those unadapted and unable to adapt.

Capitalism owed its unrivalled flexibility to the fact that market prices and earnings were an excellent means of assuring both kinds of flexibility. Price signals proved to be a quick, reliable, and accurate system of communication; the incentives they provided seemed to spur individuals to the utmost flexibility of which they were capable; and the distributive function of prices added flexibility of the second kind, because every redistribution brought about by market forces shifted the economy's center of gravity away from nonadapters, towards adapters and the adapted.

With the passage of time, however, the joints of that once wondrously flexible structure are becoming more and more calcified and rigid; and the process seems hard to reverse, because it stems from so many different and diverse causes. Growing affluence, improved technology, the greater role of government, and the increased bureaucratization and expanding size of firms are among factors that explain why individual buyers and sellers have become less responsive to market signals and tend increasingly to ignore the gains and losses with which the market rewards adaptation and punishes failure to adapt. Such a decline in people's and firms' responses to price signals would normally cause the price signals to become stronger and flexibility of the second kind to take over. However, the price signals themselves have their own rigidities, which are also on the increase due to oligopolistic and bureaucratic price fixing. Besides, the second kind of flexibility, acting through a redistribution of income, grates on our sense of distributive justice, which is why we are forever tempted to impede its action. All those factors, of course, the failure of people and firms to heed changing prices, and the failure of prices to reflect changing conditions, contribute to rendering the economy less flexible.

Let me give just a few illustrations. The most obvious example is the growth of government as a buyer of goods and services. Government usually considers itself above petty cost considerations and seldom revises its shopping list in response to changing prices. Its growing economic role therefore increases the weight of its own inflexibility.

Private firms, once so flexible and quick to exploit new opportunities and adapt to price fluctuations, have also become less responsive to changes in market conditions. One reason, probably, is the increasing complexity and size of the most efficient producing unit, due to ever more mechanized and automated productive processes; another

perhaps is the increasingly bureaucratic way in which those ever larger units are administered; and both of those factors make it harder, dearer, and slower to innovate, revise production methods, and change input coefficients. A third reason is the great importance of government contracts, which encourage the firms that receive those contracts to become just as inflexible as their main customer. Yet another reason is the ever more pervasive regulatory powers of government, prompted by concerns over health, environment, job safety, product safety, equity, and similar considerations. That last is generally considered the main impediment to the firm's flexibility; and one gets an idea of its importance from estimates of the costs it imposes on the economy. The static cost of administering and complying with government regulation is put at 5 percent of the national income. More important for us is the dynamic cost, in the sense of the slowing of innovation, adaptation, introduction of new products; and the part of that cost imposed merely by environmental and job-safety regulation is estimated at a quarter of the potential rise in labor productivity. (See M. L. Weidenbaum and Edward Denison.)

A striking illustration of what has happened to flexibility is our economy's failure to date, six long years after OPEC had started flexing its muscles, to have taken even a first step towards substituting coal for oil in response to rising oil and falling coal prices, although idle capacity, unemployed miners, and coal deposits are all plentiful. The California Energy Commission is said to need between two and one-half and three years to approve the siting and construction of a coal-fired thermo-electric generating plant; the Environmental Protection Agency took six years to grant a permit for such a plant in Montana.[1] If those cases are typical, they explain much of the inflexibility of production coefficients.

Government and business firms are, of course, complex organizations; perhaps it is inevitable that their inflexibility should increase with sheer size alone. But what about the consumer whose market behavior is a no-less important part of the behavior of the economy? Surely human nature is unchanging; how can one reconcile that fact with evidence that the consumer too has become less adaptable and less responsive to changing market prices? The obvious explanation is his increased affluence. To change one's behavior in response to worsened circumstances is, in a sense, an acceptance and admission of defeat. Not to change one's behavior is like keeping a stiff upper lip in adversity: it is manly, dignified, impressive and costly. Affluent consumers are quite able and willing to bear the private cost of such behavior; they are unaware and unheeding of the social cost, which includes the diminished flexibility and viability of the system.

One illustration is the American consumer's response, or rather lack of response, to the rising price of imports. We are fortunate enough to be able to produce at home most of the things we need and want. That is why most of our imports of consumer goods are like icing on a cake: small additions to the much larger volume of domestically produced versions of the same product, prized just for being more fancy or just for being foreign. It should have been all the easier therefore to dispense with many of them as the depreciation of the dollar raised their prices. Yet, nothing of the sort happened. As if the dollar had not depreciated, or as if price were no object, we continue to import wines, beer, shoes, automobiles, and much else besides in undiminished quantities, sometimes in undiminished proportions of our increasing total consumption of those commodities. The public seems ready to pay through the nose for the maintenance of its living standards and status symbols; but its uppity behavior robs the balance of payments of its ability to right itself, causes the dollar to depreciate further and further, and adds to our inflationary woes. When consumers don't respond to the rising cost of imports, domestic

[1]Compare the first 1979 quarterly report of the Pacific Gas & Electric Company and the *U.S. News & World Report*, Nov. 12, 1979, p. 73.

prices do; and that, of course, is the worst kind of response.

Another example of that same stiff-upper-lipmanship is the behavior of workers both here and in other advanced capitalist countries. Rising affluence, and in some countries also a temporary shortage of skills, has made them raise their sights to more important, more responsible, more prestigious jobs, and to acquire the requisite training. Unfortunately, the upgrading of the labor force ran ahead of the upgrading of jobs. The resulting excess supply of higher skills and qualifications has already depressed the relative earning power of the college trained, raised their unemployment rates, and lengthened their search periods. At the same time, the lowly, dirty, unpleasant jobs at the other end of the skill scale would go begging were it not for the large inflow of Puerto Ricans and Mexicans to this country, Pakistanis and Indians to Britain, Algerians to France, and assorted guest workers to the other affluent countries of Western Europe. By now, the number of foreign workers in the labor force of those countries is commensurate to the number of their unemployed, yet, to send them home would be no remedy, because the indigenous unemployed disdain the jobs that ousted guest workers would relinquish. They prefer, and thanks to Social Security and accumulated savings are able, to stay unemployed rather than to accept jobs beneath their station. If labor markets nevertheless retain long-run flexibility, it is owing to the next generation's response to market signals.

All those were instances of the diminished response of buyers and sellers to price signals, which leaves unimpaired the second kind of flexibility—the shifting of the economy's center of gravity. But the loss of people's and firms' responsiveness to prices is compounded by the diminished response also of prices to changing conditions; and that has not only diminished the second kind of flexibility but, by impairing the signalling and incentive functions of prices, has further reduced individual adaptation as well.

Price rigidity has many causes. Disparity between the number of buyers and sellers accounts for price setting by price makers; and the greater the disparity, the more rigidly bureaucratic the way in which price makers set their prices. For the asymmetrical downward rigidity of wage rates everybody seems to have his own explanation. Mine, of course, is by far the most plausible and sensible; but it is too long to repeat here. It has to do with the conflict between unequal incomes and equal bargaining strength on the two sides of the labor market. (See my 1978 article.) Monopolistic and oligopolistic restriction is another cause of price rigidity; and it explains, for instance, the tendency of productivity growth to lead, not to price reductions, which would be anti-inflationary, but to wage increases, which become inflationary as they are extended across the board for equity's sake, and so spread from workers whose productivity has risen to those whose productivity lags behind. Another cause of price rigidity is society's increasing refusal to accept inequitable price constellations and price adjustments.

Capitalism never got high marks for equity, although it was good not only to capitalists. Its fast growth benefited the average person and with him the majority; but it did so at the cost of great fluctuations and a wide dispersion of incomes around that fast-growing average. It provided jackpots and fleshpots for a minority, inflicted misery and unemployment or bankruptcy and ruin on another, larger minority—most of it the outcome of the great rewards and high penalties capitalism distributed in the interests of flexibility. Flexibility and the system's viability were well maintained as long as the public accepted the inequality of incomes and the wide scatter of winnings and losses as manifestations of the immutable laws of the market.

In time, however, we learned how to correct, suppress, overrule, or supplement the distributive functions of prices and price changes; and the better we learned to divorce income distribution from market forces, the more reluctant we became to accept the economic inequalities imposed by the market.

A good example is the Scandinavian countries' use of periodic centralized and synchronized wage bargains as occasions for

implementing the policy of the Trade-Union Confederation aimed at mitigating wage inequalities. Between those equitable annual wage increases, inequality sneaks back as employers raise the pay of the more highly skilled in response to their greater scarcity; and that tug of war between the forces of equity and marginal productivity is now recognized as an important inflationary factor. (See O. Aukrust.) A similar example here and in many Western European countries is the tendency for productivity-related wage increases to be extended, for fairness' sake, to groups of workers whose productivity has increased less or not at all; and that too is a well-known and much-discussed inflationary force. (See John Eatwell, John Llewellyn, and Roger Tarling.)

It is mostly government, however, that overrules market prices and suppresses competitive price adjustments. Farm price supports, airline regulation, and the dead-ending of trucks are examples of government interference in favor of business; but there is much of it also for the sake of greater equity. Indeed, to intervene in the interests of distributive justice has become an important preoccupation of government under advanced capitalism. To mitigate the inequities of capitalism while preserving its efficiency is the great liberal compromise—but does it work? Since we cannot divorce the signalling function of prices from their distributive function, every modification of market-determined distribution is likely to weaken or falsify market signalling and thereby to weaken the economy's automatic tendency to adjust. That being the case, one may well ask how the capitalist state got itself into the business of undermining its own economic base?

The answer, I think, is that up to a point the liberal compromise worked pretty well, especially while it was confined to improving the static income distribution through general taxation and the provision of social services and social security benefits. Apart from lowering the incentive to save, those measures have little or no impact on choice: the choice of consumers, workers, and producers between alternative commodities, services, inputs, and outputs. They are believed to diminish the incentive to work or to participate in economic activity; but in the beginning, they had no measurable impact on that either.

That was not only beginner's luck. It helped that, in this country certainly and in many other countries probably, progressive taxation is a myth. Our taxes make a hardly noticeable change in the distribution of income, because the main effect of the progressivity of the income tax is to undo the regressivity of the other taxes.

It also helped that price signals are not the only vehicle of market communication and the earning of money to spend is not the only incentive to work. There is satisfaction in the work itself and in one's recognition of oneself as a useful member of society. In countries with strong puritan traditions, people value income not only for what it will buy but also for the status it gives, because they regard it as a measure of their own worth to society. That is measured, of course, by income *before* taxes; and there you have the one and only instance of the successful separation of the incentive function of earnings from their distributive function.

Those probably are the main reasons why the liberal compromise was so successful in the early stages—successful enough not to seem like a compromise at all and not to cause a noticeable decline in market efficiency. It is hardly surprising that we got carried away into believing that progress towards more equality and more security could proceed almost indefinitely without impairing market performance. Doubts are arising slowly and only *ex post facto* as the blunting of market rewards and penalties is increasingly focused on assuring fairness in a dynamic sense: in the incidence of *changes* in earnings and in the distribution of the rewards and penalties created by price *changes*.

By the first of those I mean the taxation of capital gains and profits and the Scandinavian and British near-confiscatory tax rates on the very top of the income distribution. They are much blamed for reducing the rewards for risk taking and for dampening the spirit of innovation. They usually go hand in hand with government's also taking over some of the risks: allowing

tax deductions of business losses, footing the bill for *R&D*, and occasionally bailing out companies on the verge of failure. Such reductions of the entrepreneur's risk would offset reductions in his reward for taking risk, were it not confined, the first to companies large and diversified enough to have profits to offset losses against, the last to corporations so large that their failure would rock the boat. Unfortunately, the government's sharing the risks of those firms does more to ossify the corporate structure than to raise its animal spirits, while the Schumpeterian entrepreneur, the imaginative newcomer who has to risk his all to realize his vision, is left out in the cold, relieved of the reward for risk taking, but not of the risk itself.

The second form of state intervention just mentioned usually takes the form of price control proper. The United States' control of oil prices is an excellent illustration of both the depth of the conflict between equity and efficiency concerns, and the magnitude of the problems created when government suppresses market forces and puts nothing in their place.

I made inflexibility the villain of the piece but, apart from a few passing references to such things as the inflationary effects of price rigidity, and the slowing by government regulation of technical change and progress, I have said very little so far to establish its villainous nature. Let me now make good that omission by discussing a much neglected function of market flexibility: its impact on macro-economic problems and macro-economic policies.

Flexibility is the ability and willingness to substitute one thing for another, including the changing of inputs in response to a change in output and the changing of expenditures in response to a change in income. That is why, in a flexible economy, any change, disturbance, or disequilibrium originating in one market spreads to other markets, why the different disturbances in different markets are brought together and evened out, and why market situations throughout the economy tend to become more nearly uniform.

By contrast, inflexibility separates markets one from another and locks in whatever disturbance or disequilibrium each of them may be having. In other words, inflexibility causes, or rather preserves, the fragmentation of the economy, which becomes manifest in a dispersion of degrees and types of market disequilibria across sectors and regions. Empirically, that is best measured by the dispersion of unemployment rates, data on which are available, and suggest that, in the United States and the United Kingdom at least, there has been a secular increase in economic fragmentation.[2]

Such fragmentation, however, and the resulting disparities between disequilibria in different markets, are not necessarily a bad thing. After all, the diffusion of a disturbance through several markets does not usually dissipate it. In economics, when one change leads to another, a cumulative process is often generated which magnifies rather than dissipates the initial disturbance. The unemployment multiplier is the obvious example. Accordingly, there is something to be said in favor of inflexibility which locks in and isolates a fall in effective demand. Indeed, social insurance, social services, and the affluent consumer's and worker's uppity behavior, all of which I blamed earlier for impairing flexibility, also go by the name of built-in stabilizers. They have undoubtedly lowered the value of the multiplier and probably explain why these days, thank God, unemployment never reaches anywhere near the heights it used to scale in the heyday of capitalism, say, in the 1890's and 1930's.

Unfortunately, that same inflexibility is no stabilizer in the opposite direction: it fails to arrest the spreading of inflationary price increases. The reason is that inflexibility locks in the excess demand but not the

[2]Richard Lipsey seems to have first posed the question whether inflation might not also depend on the dispersion of excess demand and the consequent dispersion of unemployment across regions and sectors. His question was followed up by C. G. Archibald and by R. Leighton Thomas and P. J. M. Stoney. They both found a significant and quantitatively important positive relationship; and Archibald explained the upward shift of the Phillips curve by increased dispersion. For a qualification, see Archibald, Robyn Kemis, and J. W. Perkins. My argument, of course, is quite different, but clearly related.

resulting rise in prices, because excess demand is not the only carrier of the inflation virus. Wage increases, for example, spread from firm to firm and industry to industry by the principle of equity more than *via* the spreading of labor shortages—and the latter alone is prevented by the fragmentation of markets.

Inflexibility shows up in even worse light in the next case. When two markets or two parts of the economy suffer, one from an excess, the other from an insufficiency of demand, flexibility would enable those two contrary disequilibria mutually to offset and eliminate one another, inflexibility preserves both of them intact. That may explain why we have more economic troubles nowadays than there used to be in the days of flexible capitalism, and also why today they seem so much more intractable.

That brings me to the next and by far the most serious trouble that inflexibility creates: its tendency to render orthodox Keynesian stabilization policies ineffective or politically unacceptable. I mentioned that flexibility and interaction between markets integrates their disequilibria, offsets one against the other, diffuses them, and renders them more uniform. That opens the way to cumulative processes; but it also has the inestimable advantage of greatly facilitating the use of stabilization policies by so-to-speak preparing the patient for the treatment.

Fiscal and monetary restriction and expansion are macro-economic policies par excellence. They cannot be focused on only one sore spot in the body economic; their influence is broad gauged, diffused, and so acts on the economy's different sectors and industries more or less equally and with the same intensity. That is why they are the most effective and politically the most acceptable in an economy whose different parts and sectors all suffer from the same trouble and more or less to the same degree. Flexibility, of course, is what creates that kind of uniformity, because it leads to high elasticities of substitution, transformation, adaptation, and thereby assures that the different parts of an underheated or overheated economy will have not too dissimilar temperatures.

That may well explain why, in the more flexible and therefore more uniform economic environment of a generation ago, when stabilization policies were first developed and applied, they were effective and acceptable; whereas in today's rigid, fragmented, and heterogeneous economy, they often seem both ineffective and unacceptable. Indeed, what can fiscal and monetary policy do against the kind of stagflation where insufficient demand in one part of the economy creates unemployment, and excess demand in the other part creates inflationary pressures, which however do not remain confined to that part alone. What can be done even in the much simpler case where one part of the economy is overheated and the rest more or less in equilibrium? You get inflationary pressures generated in the overheated part and spreading out from it to the rest of the economy; but try to contain them with restrictive macro policies and you are sure to create unemployment much sooner than you manage to slow the rise in prices. Policymakers complain that the public has become oversensitive to the side effects of anti-inflationary policies: is it not rather that the side effects have become more severe, now that those policies are applied in a more fragmented, more heterogeneous economy than they were designed for? Perhaps the most ominous consequence of inflexibility is that by failing to homogenize the economy, it limits the effectiveness and acceptability of macro-economic policies.

The relevance of all that for the question to which this lecture is addressed is obvious. If I assessed the role and the fate of flexibility correctly, then I have added strength to Schumpeter's argument. Capitalism is bound to lose drive and manageability if it is becoming less flexible and less adaptable.

Lest you find that thought too depressing, let me remind you that Schumpeter predicted the demise of capitalism at the very time when it was embarking on its longest stretch of uninterrupted fast growth and high prosperity. Had he known that, he might still have stuck to his prediction, which was based on sociological more than economic considerations; but I do believe that he failed to recognize the importance of

Keynes' contribution and failed to foresee the success of Keynesian macro policies in stabilizing the free-enterprise economy without making it less free and less enterprising. One of my main points is that today's economy is too fragmented for the broad-gauged, diffused action of Keynesian policies to do much good; but perhaps I, too, am overlooking the emergence of a new way out.

Salvation could lie in recapturing some of that lost flexibility or in some new policy better able to deal with today's problems. We should really have both. But lost flexibility is hard to restore, because some of it was lost due to factors beyond our control, the rest was the side effect of liberal policies, ignored or minimized at the time owing to the liberals' excessive faith in capitalism and its ability to fly, however much its wings are clipped. It is difficult to go back on those policies, because they secured important benefits, which the public is reluctant to give up, whether or not they are worth the cost.

Yet, difficult as it may be to recapture lost flexibility, we have taken a few steps in that direction already. One of them is the budget constraints California's Proposition 13 imposed on government and not only in California. Another one is airline deregulation; and several other of our regulatory agencies are also under attack. A different example, but perhaps only to show that the old flexibility is not all dead, is the competitive pressure thanks to which progress in electronics has led to stabilizing price reductions instead of inflationary wage increases.

Now what about new policies to deal with our new problems, and our old problems in their new setting? Nothing so far approximates what happened in the depressed 1930's: the spreading like wildfire of Keynes' new ideas and their capturing the imagination and loyalty of an entire generation of young economists. I know of some good new ideas and there must be many more; but none has caught on. One reason could be that to deal effectively with today's economic problems, new policies may have to be too selective or too drastic to stay within the spirit of capitalism. The benefits could well be worth the costs; but that unfortunately is not enough, because the public, jealous of its accustomed freedoms, may refuse to give such policies a fair hearing.

For this country's faith in capitalism is very peculiar. We blithely pile regulation on regulation, imposing as many costs and fetters upon the functioning of our markets, with never a thought of what that may do to their efficiency; and at the same time, we retain more faith even in the most hamstrung and ineffectual markets and market signals than in anything that would displace or bolster them or change their character. The danger with such an attitude is that we might end up with the worst of both worlds: stuck with capitalism's unattractive features, but deprived of what I described at the outset as one of the system's great and redeeming virtues: its flexibility.

REFERENCES

C. G. Archibald, "Wage Price Dynamics, Inflation, and Unemployment, the Phillips Curve and the Distribution of Unemployment," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 124–60.

______, **R. Kemis, and J. W. Perkins**, "Excess Demand for Labour, Unemployment, and the Phillips Curve: A Theoretical and Empirical Study," in David Laidler and D. L. Purdy, eds., *Inflation and Labour Markets*, Manchester 1974, 109–63.

O. Aukrust, "Inflation in the Open Economy: A Norwegian Model," in L. B. Krause and W. S. Salant, eds., *Worldwide Inflation—Theory and Recent Experience*, Washington 1977.

J. Eatwell, J. Llewellyn, and R. Tarlin, "Money Wage Inflation in Industrial Countries," *Rev. Econ. Stud.*, Oct. 1974, *41*, 515–23.

E. F. Denison, "Effects of Selected Changes in the Institutional and Human Environment Upon Output per Unit of Input," *Surv. Curr. Bus.*, Jan. 1978, *58*, 21–44.

R. G. Lipsey, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: A Further Analysis,"

Economica, Feb. 1960, *27*, 1–31.

T. Scitovsky, "Asymmetries in Economics," *Scottish J. Polit. Econ.*, Nov. 1978, *25*, 227–37.

R. L. Thomas and P. J. M. Stoney, "Unemployment Dispersion as a Determinant of Wage Inflation in the U.K. 1925–66," *Manchester Sch. Econ. Soc. Stud.*, June 1971, *39*, 83–116.

M. L. Weidenbaum, "The Big Cost of Government Regulation," *Challenge*, Nov./Dec. 1979, *22*, 32–39.

Post-Keynesian Economics: A Promise that Bounced?

By L. TARSHIS*

Any golden-aged economist who heard his or her elders criticize John Maynard Keynes' *General Theory* shortly after its publication has learned to be wary lest he or she in turn now sound like them. "It's all been said before, and what is new is wrong" was their theme. Yet if what Keynes wrote was either old or wrong, my reading of today's versions of the neoclassical synthesis persuades me that they are old and wrong and, in addition, no less dangerous than they were in 1930 or 1936! My first point of agreement with Post-Keynesian economics then is that it is on the right track in criticizing the neoclassical paradigm.

It has other virtues too. It embodies an attempt to extend, clarify, and bring up to date Keynes' *General Theory*: a task which Keynes himself intended to undertake, as he promised Ralph Hawtrey less than six months after it was published. It aims to develop the micro foundations of macroeconomic theory. It finds guidance in two decidedly important features of Keynes' thought which had been almost obscured by the neoclassical synthesis: that economic processes occur in *historical* rather than in *logical time*; and that practically all decisions at firm, household, and government levels must be made in the face of serious uncertainty. It is concerned, moreover, with several problems which stem from Keynes' work but which he did not tackle: for example, the theory of the distribution of the national income, and the theory of economic growth and its accompanying fluctuations. In short, it has set for those who work under its banner an impressive agenda. But the question still remains as to whether its achievements come at all close to matching the challenge. With a few notable exceptions, its ambitious claims seem empty of substance.

*Director of Research, Ontario Economic Council, and professor emeritus, Stanford University and the University of Toronto.

Let us look, to begin with, at the very base of the micro foundations which Post-Keynesians use—I mean the analysis of price and output decisions. Their model of the economy consists of a central core of oligopolistic firms—I am tempted to call them "multinational, oligopolistic megacorps" to preserve the flavor of the analysis—which administer prices. They are surrounded by a ring of smaller firms, chiefly in the raw materials, agricultural, and light manufacturing sectors whose prices are established by the Invisible Hand. But their account of price setting in the central core strikes me as either saying nothing, or as being wrong, with unfortunate consequences for much that follows.

My analysis begins with the observation that each firm in the central sector sets price by adding a certain markup to its average *variable* costs (AVC). Thus $P = AVC + (m \times AVC)$ where m is the markup ratio and $(m \times AVC)$ is the actual markup in dollars. Now clearly in "normal circumstances," markup pricing *even with an unchanged markup*, can yield a price which is very close to the one the firm would obtain if it were to produce as much as would bring its marginal revenue (MR) to equality with its marginal costs (MC) and then to set, for the whole output, the highest price it can get. Whether its profits would in fact then be at their maximum figure, or less, would depend upon its skill in choosing the proper markup.

This result holds in its simplest form provided that the firm's AVC is invariant to

changes in its output, and there is plenty of evidence to confirm that this condition is realistic enough.

When AVC does not change with output over a certain range, the marginal cost of any output within that range is the same as AVC. Hence we may substitute AVC for MC.

A second definitional identity has $MR = P \times (e-1)/e$ where P is price and e is the elasticity of the demand for the product.

For profit to be at a maximum, MR must equal MC. Replacing MR by $P \times (e-1)/e$ and MC by AVC we have that $P \times (e-1)/e$ must equal AVC. Hence P must equal $AVC \times e/(e-1)$ which can be simplified to $P = AVC + AVC/(e-1)$. This is simply AVC plus a markup on AVC which in percentage terms is $100/(e-1)$.

It follows that if the markup is properly chosen—say by a Darwinian process in which firms that have somehow made a near-correct choice survive and hence can instruct newcomers—and if changes in the value of e are generally small, markup pricing with a *constant* markup will yield a price close to "the most profitable one." And the rule is obviously far easier to use!

When the rule for the profit-maximizing markup is formulated so as to take account of the elasticity of the AVC (η) as well as the demand elasticity (e), it becomes: $m = (e+\eta)/(e-1)\eta$. The lower is η for any given e, the larger has to be the markup for maximizing profit.

Accordingly, when the firm is trying to maximize its short-run profits, and when it must therefore expect the AVC to become less elastic as its output approaches capacity, the size of the most profitable markup grow enormously, yet there is no evidence that it does. But when we assume that the firm aims at maximizing its *long-run profits*, then since η will remain near constant over a much wider range, the near-constant markup once again makes sense. Obviously, however, the markup would not be fixed over the whole range. A few increases would have to be anticipated.

All this is to stress that there is no inherent contradiction between markup pricing with a conventional markup—a practice that we are told price-making firms generally follow—and the profit-maximizing assumption that conventional economics accepts.

But for reasons which remain obscure some Post-Keynesian economists assume that firms in the crucial core are part-way philanthropists; that they don't even aim to maximize profits. Instead, according to this view, each firm seeks only to maintain or possibly to increase its share of the (growing) market. To do this, a certain level of investment spending is called for, and the markup is set at the level that will generate enough undistributed profits so that along with its allowance for capital consumption it can finance all or at least most of its investment outlays internally.

I believe this view is not only in conflict with the firm's behavior (there is no evidence that markups change so often), but it is also far too simpleminded. It seems to rest upon the observation that in most years gross corporate saving has ranged from 70 to 90 percent of corporate outlays for investment projects; which has led some Post-Keynesians to suppose that individual firms are largely self-financed. But to conclude that generally the firm pays for, say, three-quarters of its investment outlays with internally generated funds, simply because we find that the whole sector saves as much as three-quarters or nine-tenths of its gross investment outlays is an unwarranted leap. In fact, in 1978 nonfinancial corporations raised over $130 billion of long- and short-term funds externally, and their total reliance upon external finance amounted to at least 55 percent of their investment outlays. Though the corporate sector as a unit did not have to rely upon externally raised funds for most of its investment outlays, it seems clear that many firms were much less close to financial self-sufficiency than Post-Keynesian micro theory implies, while others, of course, must have had a surplus.

There are additional reasons for doubting that the typical firm determines its markup so as to yield enough to enable it to finance most of its investment outlays internally.

First of all, firms often prefer to borrow instead of increasing equity. Admittedly an

increase in debt would be unwelcome to a firm with too high a debt-equity ratio and such a firm, its undistributed gross earnings permitting, might even use the counterpart funds to acquire other assets or to repay debt. But generally firms do not show such an aversion to debt. Instead, they anticipate the gains that increased leverage permits and are quite ready to have their debt-equity ratio rise.

Next, even though firms do not want to borrow to finance investment, they may find that raising markups does not increase their profits. After all, if the firm has already set its markup and prices so as to yield maximum profits, an increase in its markup would cause its profits to fall. Moreover, if firms are oligopolists, with each uncertain as to how its competitors will respond to any price initiative it takes, they are likely to hold their prices fixed against small shocks; in particular, each may be unready to raise a too-low markup because of its fear that its competitors would leave their prices unchanged and thus jeopardize its profits. Incidentally, it is hard to reconcile the Post-Keynesians' insistence that uncertainty about the future is ever present, with their readiness to forget this whole business when they come to discuss price setting in that very set of market conditions in which uncertainty is likely to be most serious. Such a fault in their explanation of price determination will surely upset theories of distribution and of investment outlays derived from the basic theory.

Let us consider first the Post-Keynesian explanation of investment spending. With investment outlays mainly financed internally, money may be supposed to play only a minor role; after all, when a firm does not need to borrow from the market its sensitivity to the cost of borrowing is likely to be low. Some Post-Keynesian economists have concluded that if interest rates play any role at all in connection with investment spending, they do so by influencing the manner in which funds are raised rather than by affecting the total amount firms will spend to finance investment projects. These views are supported by a number of "empirical investigations" which seem to show that the individual firm's investment decision is not sensitive to changes in interest rates. And from this observation, they have concluded that the investment outlays of the whole economy are also interest inelastic. But this conclusion does not follow at all.

This can be seen most readily if we suppose that each firm's demand for investment goods is completely interest inelastic, except at one point. Firm *A* will order ten units of investment goods so long as the interest rate is anywhere below 20 percent, and none at all if it goes above. For Firm *B*, with another ten units in question, the cutoff rate is 18 percent; for *C* it is 16 percent; for *D*, 14 percent; for *E*, 12 percent; and for *F*, 10 percent. Then if the rate were 9.9 percent, every firm would order its full quota of investment goods—or sixty in all. An increase in the rate to 11.9 percent would leave five firms ordering fifty units *in toto*; at 13.9 percent, four firms ordering forty units, etc. Their aggregate demand for investment would be shown, apart from discontinuities, as a straight line with a slope such that a change of one percentage point would lead to a change in the opposite direction of five units; the elasticity of the function at mid-range would be approximately 2.5. An even more extreme conclusion would be reached if all firms were assumed to have nearly the same cutoff point (most unlikely!). Notice too that each of these strikingly different results is consistent with a completely inelastic demand for investment by the individual firms. Note also that the only way in which the elasticity of the function over a wide range of interest rates could be zero would be if cutoff rates were very dispersed: say 25, 24, 4, 3, 2, and 1 percent. Then between 23.9 and 4 percent, the demand would be completely inelastic. Whether the firm's demand for investment is interest inelastic or not seems to have very little effect on the investment function for the whole economy.

With a few exceptions, Post-Keynesian theories of investment are no less flawed and incomplete than Keynes' own analysis. Two points in particular bother me. For one thing, their piety too often leads them to invoke "animal spirits" as described by

Keynes, as a sufficient explanation of changes in investment outlays, and that, it seems to me, is no better than to sweep the problem under the rug. But perhaps economics can go no further. My second objection is that too often they fall into the trap which I had supposed Keynes' *General Theory* was to remove; that an act of saving by one firm would lead the saver to increase its and the economy's investment outlays. But such a result can only be guaranteed if the levels of income and output are fixed.

Finally, let us turn to Post-Keynesian theories of distribution. One tradition which stems from Nicholas Kaldor has been to explain the functional distribution of income, and specifically the share of profits in national income, as equal to the share of investment in national output. This result depends upon the assumption that workers save nothing, spending the whole of their income to purchase consumer goods, while employers consume nothing, instead saving all they earn. With more realistic assumptions, the results must be modified, but the dominant role of the ratio I/Y remains. This whole analysis rests or at least seems to rest upon *demand factors*—purchases of investment goods, and perhaps also the difference between workers' and employers' marginal propensities to consume. And because it rests wholly upon aspects of demand, it is subject to the type of criticism Post-Keynesian economists often direct toward simpleminded Keynesian doctrines. What is more, as we shall see, its "causal variables" are really endogenous.

There is, however, a second tradition, or a second basis for a macro theory of distribution, which Post-Keynesian economists occasionally use. It rests upon aspects of *supply* and especially upon the relation between costs and prices inside the various firm(s).

Consider first, short-run changes in the distribution of income. If wages account for most of the economy's variable costs of production—and except for imported raw materials and certain taxes, this is a reasonably good approximation to reality—and if average variable costs are, within a wide range of output, more or less constant, it then follows that the ratio of wage costs (and hence wage incomes) to the aggregate receipts of all firms, after netting for any double counting, will remain approximately the same, when the aggregate output of the economy changes. (This, of course, assumes that each price is set at a fixed markup over average variable costs.) The ratio of nonwage income (profits *plus* various components of fixed costs) will then also be unaffected, or nearly so, when output changes within that range. But if output is pushed into the range in which the average variable cost curve is sloped upwards to the right, we must face a possible discrepancy between the two traditions. Going back: if the ratio S/Y is constant, the ratio I/Y and profits to income will also be constant when I is increased. Hence it follows labor's share of the national income would not change over the whole range. But if an increase in investment were to push the outputs of most firms into the region in which their AVC rose with such increases, the ratio of the marginal cost of any output to its AVC would also rise with increases in output. Then even with no change in the elasticity of demand for the product the appropriate (the *profit-maximizing*) percentage markup over average variable costs would then also rise as output rose. With the markup over the AVC now higher than it has been, labor's share of the national income would fall as output grew. But this result is at odds with that obtained by looking at the ratio I/Y. Post-Keynesian economists seem to use both these approaches, but, so far as I have seen, they use them as though they were identical.

This paradox is resolved when it is noted that if we grant the assumption that workers save nothing and nonworkers save all they earn, any change in the functional distribution of income will affect the ratio S/Y, i.e., the saving function. When output is expanded into the range in which AVC rises with output, the ratio of profits to income will rise also (once markups increase). Even though the saving habits of workers and nonworkers considered separately do not change, the average propensity to save of the whole economy must also rise. Hence at

each income level in that range, the ratio of S/Y and hence I/Y will be higher than it would have been if these ratios were taken as exogenous and constant. But note, this puts the explanation for any change in the functional distribution of income where it belongs: *with costs and markups*, not with the ratio of S/Y or I/Y. The ratio S/Y is determined by the distribution of income.

This consideration also helps to broaden the knife-edge that plagues Harrod's growth model. When the actual growth rate exceeds the warranted rate, and the economy hence faces an explosion, the rise in the *saving ratio*, which follows if we accept these assumptions, will move the warranted rate towards, and hopefully right to the actual rate; we then have something like a fail-safe mechanism to secure a stable growth rate. But as we should by now realize the assumptions about the savings propensities for these two classes of income recipients are dubious.

It is unfair, I admit, to fault Post-Keynesian economics for its omissions. What little revolution in thought it embodies, is far from complete. But the sympathetic viewer from the sidelines may at least express the hope that its members will devote attention to some of the problems not yet tackled. Two, in my judgement, are prominent.

a) Hyman Minsky's contributions to our understanding of the processes through which growth may lead to financial crisis, though presumably written in the Post-Keynesian tradition, differs from that literature in assuming that the leading firms are willing to use debt financing when their investment outlays are high instead of financing most of their investment internally in a period of growth. I believe, for reasons set out above, that Minsky's insight should be incorporated into the rest of Post-Keynesian economics and this would clearly call for a revision in their model of how investment is financed.

b) It has been usual in dealing with macro-economics problems to assume that the market economy by its day-to-day adjustments handles the problems of resource allocation more or less adequately. As a result very little attention has been paid to the possibility that a failure of the micro-adjustment process could affect adversely the level of aggregate output or the price level, or that such a failure could hamper macro-economic policies. Now it is doubtful whether such an assumption was appropriate for Britain in the 1930's: efforts to support the persistently slumping textile or shipbuilding industries or the coal mines suggest that at least a part of Britain's unemployment problem in, say, 1931, reflected a secular maladjustment which would not disappear by a single turn of the demand lever to full expansion.

It is even more doubtful that their assumption is appropriate today. The refeudalization of capitalism has proceeded quickly—in the United States and Canada certainly. This process both reflects and validates the expectations of the various groups able to protect their share of the national income. Insofar as they succeed, the economy's ability to adjust to shocks is likely to be weakened. Prices will show less "give" especially in the downwards direction, and macro policies are less likely to work. Macro-economic theory should be updated to allow for these inflexibilities, possibly by restoring Keynes' forgotten aggregate supply function.

On Keynesian Economics and the Economics of the Post-Keynesians

By JANET L. YELLEN*

This paper compares the new Post-Keynesian economics of Paul Davidson, Alfred Eichner, Geoffrey Harcourt, Jan Kregel, Hyman Minsky, Sidney Weintraub, and others with the standard Keynesian macro-economic model which these authors have criticized. Sections I–IV concern, respectively, Post-Keynesian views on pricing, output and income distribution, inflation, and money and finance.

I. Post-Keynesian Price Theory

According to Post-Keynesians, competitive firms for which price equals marginal cost are the exception rather than the rule. Post-Keynesians point to the prevalence of "megacorps" and "administered prices," concluding that realism demands a scrapping of the perfect competition paradigm. Following Michal Kalecki and Joan Robinson, the progenitors of their movement, most Post-Keynesians assume that firms' prices are a markup over unit production cost. Many Post-Keynesians have argued that prices are insensitive to demand fluctuations but sensitive to permanent cost changes, one explanation being that many large firms have marginal and average variable costs which are virtually constant over large ranges of output. Since profit-maximizing monopolists set price equal to marginal cost times a factor which depends on the elasticity of demand, this explanation makes sense if the perceived elasticity of demand also remains constant over the business cycle. An alternative hypothesis is that firms simply engage in "normal cost pricing" and their pricing decisions ignore temporary changes in unit costs or demand elasticities. Despite Post-Keynesians' claim that their view of the pricing decision is in sharp conflict with standard analysis, empirical evidence on normal pricing has been provided by such Keynesians as Otto Eckstein, Robert Gordon, William Nordhaus, Arthur Okun, and George Perry.

In contrast, Eichner's view is novel. According to Eichner, the external cost of borrowing exceeds the cost of using retained earnings. Firms that set prices below the short-run-maximizing level (as a deterrent to entry or in fear of government intervention) will tend to raise their prices in the short run to finance profitable investments. The next section explores the macro-economic implications of Post-Keynesian pricing.

II. Distributional Effects and Aggregate Demand

The Post-Keynesian theory of output, employment, and income distribution starts from the assumption that the average savings propensities of workers and capitalists (or alternatively, out of wage and profit income) denoted s_w and s_p differ. Just as in the *IS-LM* model, the Post-Keynesian model requires as a condition of short-run equilibrium that desired expenditures be equal to output, or equivalently, that savings equal investment. Formally, this requires that

$$(1) \qquad s_w\left(\frac{wN}{P}\right)+s_p\left(Y-\frac{wN}{P}\right)=I$$

where w and P denote the money wage rate and price level, N and Y denote employment and real output, and I denotes real investment. In most Post-Keynesian work, investment is treated as exogenous, said to be governed by "animal spirits." An alternative hypothesis is that $I/Y=\alpha$, and the rate of investment α is exogenous. Naturally,

*Lecturer, London School of Economics.

employment is a function of output produced. Most Keynesians would assume diminishing marginal productivity of labor. It is common in Post-Keynesian models, however, to assume a proportional relationship of the form $N = nY$, which follows from the assumption of a fixed-coefficients technology.

Equation (1) is the Post-Keynesian equivalent of the *IS* curve. Instead of giving alternative values of output and the interest rate consistent with equilibrium in the commodity market, it gives alternative combinations of output and the real wage consistent with commodity market equilibrium. Taking I as exogenous and assuming $s_w < s_p$, an increase in the real wage raises aggregate demand and output; hence the Post-Keynesian *IS* curve relating Y and w/P, which is depicted in Figure 1, is upward sloping. Under the alternative hypothesis that $I/Y = \alpha$, however, equation (1) can be reduced to

$$\text{(2)} \qquad \frac{w}{P} = \frac{s_p - \alpha}{n(s_p - s_w)}$$

The *IS* curve in this case is horizontal. This conclusion needs to be modified of course when there is other autonomous spending or when n depends on Y.

Taken alone, equation (1) is an incomplete model which determines nothing at all, although some Post-Keynesian literature suggests otherwise. Post-Keynesians close this model in a variety of ways, some of which I believe to be inconsistent and so will discuss the various alternatives briefly.

One way of closing the Post-Keynesian model is by specifying that output is set exogenously at some level, $Y = \bar{Y}$. The result is the Kaldor model. One possible rationale for this assumption is that, with fixed coefficients technology, firms would want to hire labor in sufficient quantity to fully utilize existing capital as long as the real wage is no greater than the average product of labor, $1/n$. In essence, one adds to Figure 1 a second curve called *MM*, which relates the output firms would choose to supply to the real wage they face. Implicit in the Kaldor model is a kinked *MM* curve, indicated by the broken line MM_1 in Figure

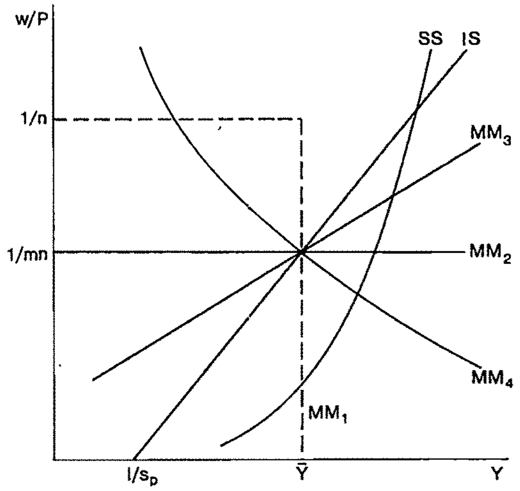


FIGURE 1

1. As long as the real wage which satisfies (1) is less than $1/n$, capital, if not labor, will be fully employed. With a fixed money wage rate, this equilibrium can be achieved via price flexibility with output prices rising whenever aggregate demand exceeds $\bar{Y}$ and vice versa. The Kaldor model has the characteristic emphasized by most Post-Keynesian writers that the level of profits, denoted Π, and the share of profits in output are directly related to I or I/Y:

$$\text{(3)} \qquad \frac{\Pi}{Y} = \frac{\frac{I}{Y} - s_w}{s_p - s_w}$$

In the extreme case which Post-Keynesians are fond of analyzing in which $s_w = 0$ and $s_p = 1$, $\Pi = I$ and "capitalists get what they spend while workers spend what they get." If $s_w = 0$ and $s_p < 1$, $\Pi = I/s_p$ and the more capitalists spend, that is, the lower s_p, the higher the profits earned.

A second way of closing the Post-Keynesian model which is more Keynesian in spirit, since it permits outcomes with less than full employment, is to drop the Kaldorian flexible markup assumption and instead postulate a fixed markup m of prices over unit labor cost so that $P = mwn$ or $w/P = 1/mn$. In this case, the *MM* curve, indicated by the solid line MM_2 in Figure 1

is horizontal. Here, firms' pricing policies determine the real wage and share of output accruing to workers and the multiplier determines the level of output as follows:

$$(4) \qquad Y = I/(s_p - (s_p - s_w)/m)$$

The denominator of (4) is the economy's average propensity to save and an increase in investment raises output by the standard Keynesian multiplier, while an increase in the markup lowers output and the real wage. Again, profits are positively related to investment but for the usual Keynesian reason that an expansion of output with a fixed profit margin augments total profits.

The postulation of a fixed markup and a fixed level of investment produces a consistent model. Suppose, however, that $I/Y = \alpha$. This small switch in assumption produces a model which, although widely discussed, is nevertheless inconsistent. With proportional savings propensities, constant labor productivity and no autonomous spending other than investment, the *IS* curve is horizontal. If the markup is a constant determined by the "degree of monopoly," the *MM* curve is horizontal too, and output is indeterminate while the real wage is overdetermined. Post-Keynesians are fond of stating that (3) determines Π/Y as a function of I/Y and that this holds regardless of how Y is determined. But this is clearly incorrect, for if the markup is fixed and the resulting real wage differs from that given by (2) the model has no solution at all. This problem is responsible for a great deal of confusion by Weintraub, Kregel, and others.

A variant of the fixed markup model is the target return pricing model which relates the markup to "normal" output. The markup may be chosen to yield a target rate of return on capital of ρ^* or total profits of $\Pi^* = \rho^* K$ at a stanard operating rate Y^n. In this case, the required markup depends inversely on normal output.

$$(5) \qquad \Pi^* = \left(1 - \frac{wn}{P}\right) Y^n = \left(1 - \frac{1}{m}\right) Y^n$$

Firms may leave m unchanged in the face of transitory changes in Y, but in equilibrium, if output expectations are satisfied we have that $Y^n = Y$. The combinations of Y and w/P which satisfy (5) for a given profit target lie on a positively sloping line such as MM_3 in Figure 1, which can be shown to be flatter than the *IS* curve unless $s_w = 0$, in which case the curves either coincide everywhere or nowhere and the model is again inconsistent. If $s_w = 0$, profits are identical at all Y, w/P combinations on the *IS* curve. If entrepreneurs are satisfied with this profit level, output is indeterminate and if not, their demands cannot be satisfied. Even if $s_w > 0$, the model has no equilibrium if entrepreneurs are too greedy. Maximum profits of $\Pi = I/s_p$ occur at $w/P = 0$ on the *IS* curve. If the firm tries to generate profits in excess of I/s_p, the model is again inconsistent. In terms of Figure 1, the relevant MM_3 lies everywhere below the *IS* curve in the positive quadrant.[1]

When well behaved, the model with target return pricing differs significantly from the standard Keynesian model with perfect competition. Under target return pricing, an increase in investment raises output but it also lowers the markup or raises the real wage. In contrast, an increase in investment raises output but lowers the real wage in the textbook model. The difference occurs because the textbook model assumes smooth factor substitutability and perfect competition or marginal cost pricing so that the supply of output and demand for labor are inversely related to the real wage. The result is an *MM* curve like MM_4 in Figure 1 which is smoothly downward sloping and is the neoclassical analogue of the kinked MM_1 obtained from fixed coefficients. Alternatively, the aggregate supply curve,[2] which is

[1] If we assume, following Eichner's suggestion, that firms wish to finance investment solely by retained earnings and interpret $s_p\Pi$ as retained earnings, then retained earnings are equal to investment at all points on the *IS* curve if workers save nothing and only at the point $w/P = 0$, $Y = I/s_p$ if $s_w > 0$ and there is no other autonomous spending.

[2] To avoid possible confusion it is important to note that the terms "aggregate supply curve" and "aggregate demand curve" refer to standard textbook relations between real output and the price level, in contrast to the alternative Post-Keynesian usage of these same terms.

assumed by Keynesians to be upward sloping, is negatively sloping under target return pricing.

Further variations on the themes outlined above are possible. Following Eichner's suggestion, the desired markup may depend on the level of investment so that an increase in investment shifts both the *IS* and *MM* curves down in Figure 1. Depending on how large the response of the markup is to a change in investment, a variety of outcomes are possible. In analyzing a similar model, Harcourt has shown that an increase in investment can cause output to fall if the effect on the markup is sufficiently strong. But if firms try to augment profits by an amount no larger than the change in desired investment, this perverse output effect is ruled out. The model could be dynamized by postulating a relationship between investment and lagged profits or sales à la Kalecki. The resulting model would be a variant of the usual multiplier-accelerator model and the cyclical possibilities inherent in such a mechanism are obvious.

III. Labor Supply, Inflation, and the Money Wage Rate

One thing notably absent from the Post-Keynesian model is a labor supply function. The *SS* curve in Figure 1 shows the output which would be produced if all labor wishing to work at a given wage were hired. It is upward sloping because higher real wages raise labor supply so that more output can be produced. There is no reason for the intersection of the *IS* and *MM* curves to coincide with the full-employment equilibrium, given by the intersection of the *MM* and *SS* curves, unless further considerations are brought into play.

In the neoclassical model, all markets are assumed to clear continuously and the real wage is determined in the labor market, that is, by the intersection of the *SS* and *MM* curves. Monetary adjustments move the *IS* curve to their mutual intersection point. According to the textbook Keynesian model, the labor market may be out of equilibrium; output and the real wage are determined by the *IS* and *MM* curves in the short run, although monetary considerations also play a role via the standard *LM* curve. But in this model, a money wage rate sufficiently low will permit a full neoclassical equilibrium. (This is the consequence of the neoclassical synthesis.) In the standard Keynesian model, money wages and hence prices change when the economy is away from full equilibrium. Of course, the full-employment equilibrium may be unstable or the path to it, even when stable, may be slow and tortuous.

In contrast, according to Post-Keynesians, the real wage and output are determined in the commodity market—at the *IS-MM* intersection—and no mechanism exists in the Post-Keynesian model which can bring the system to a full equilibrium. Post-Keynesians typically argue that labor cannot determine the real wage via money wage bargains because the commodity market determines a unique market-clearing real wage and money wage cuts just lead to proportional price changes.

Labor supply behavior does matter to Post-Keynesians in determining money wages and, in turn, inflation. They argue that money wage bargains are largely unpredictable and are made infrequently, so they should be modelled as exogenous, a sentiment shared by many standard Keynesians, being the motive for the textbook "fixed" or "predetermined" money wage model. In the Post-Keynesian model, the mood of labor and historical and institutional factors play a role when powerful unions bargain with powerful corporations. Nevertheless, Post-Keynesians believe that a wage-price spiral will ensue whenever labor regards the real wage as "unacceptably low." Inflation is a struggle over the appropriate distribution of income and it can be set off by factors which tend to increase the markup. Eichner stresses that inflation will occur if firms raise their markups to finance higher investment.

In the Post-Keynesian model, higher money wage demands necessarily lead to higher prices without raising the real wage and so there is nothing to stop inflation once it begins. Why inflation in these circumstances should not accelerate continu-

ously I cannot understand. And since the real wage which is determined by the commodity market is only acceptable to labor by accident, it is hard to fathom why such an economy should not perpetually experience either inflation or deflation.

Post-Keynesians conclude from their analysis of inflation that the solution lies in incomes policy. What is needed is societal agreement concerning the appropriate distribution of income. This policy conclusion is supported by all Post-Keynesians, and Weintraub in particular has done pioneering work on tax-based incomes policies. With markup pricing, taxes might be passed along into prices, so if the Post-Keynesian model is correct, the policy could well be pro- rather than anti-inflationary. These ideas are innovative and worthy of exploration, although the Post-Keynesian gap here between theory and policy is quite large.

IV. Money and Finance

Although Minsky and Davidson have discussed monetary factors at length and have developed a model resembling Tobin's "q" theory of investment, no Post-Keynesian has shown how money should be incorporated into their model of distribution and growth. But the inclusion of money in the model described above makes it all but indistinguishable from standard Keynesian theory. Add to the model of Figure 1 the standard LM curve, some interest sensitivity of investment or saving and possibly some wealth effects and what is obtained? The IS curve now depends on the interest rate as well as the real wage and together with the LM curve it determines Y and r corresponding to any price level and money wage. Vary the price level, holding the money wage fixed to trace out a standard downward-sloping aggregate demand curve. A cut in P raises w/P and reduces saving; but a cut in P also raises M/P, lowering r and raising I and induces wealth effects which may stimulate consumption. The MM curve relates prices and output for a fixed money wage, so it is just the usual aggregate supply curve in disguise. Under fixed markups, the aggregate supply curve is flat instead of upward sloping; under target return pricing the slope is *negative*. Together, these pieces determine short-run equilibrium values for Y, r, and P. Are there values of the money wage and price level which permit the full-employment equilibrium to be attained? The answer is clearly affirmative, for with monetary factors included, there must exist a price level sufficiently low that the full-employment output can be sold at the full-employment real wage. This is the conclusion reached by the neoclassical synthesis and there is nothing in the amended Post-Keynesian model to deprive that synthesis of its validity, however much they may despise it. Will a cut in money wages in the presence of unemployment raise the equilibrium levels of employment and output? As Weintraub stresses, a cut in w shifts *both* the aggregate supply and demand curves down, because any change in the money wage influences saving. But it is straightforward to show that the new equilibrium occurs at a higher output level as long as the slope of the aggregate supply curve exceeds that of the aggregate demand curve.

Various perversities can, of course, arise under target return pricing. In the unlikely event that the aggregate supply curve is steeper than the aggregate demand curve a cut in the money wage will *lower* equilibrium output, so the full-employment equilibrium is likely to be unstable. Furthermore, if the implicit labor demand curve MM is steeper than the labor supply curve SS, money wages may rise rather than fall when output and employment are below their full-employment levels. With respect to stability and disequilibrium behavior, the Post-Keynesian model could differ significantly from the standard textbook case, at least under certain conditions. Just what these conditions are however, we do not yet know, because Post-Keynesians have argued that events take place in "historical" rather than "logical" time and therefore have been unwilling to conduct the standard dynamic analysis.

Post-Keynesian Economic Theory: An Overview and Evaluation

By JAMES R. CROTTY*

The degree to which the economics profession is intellectually and emotionally committed to the prevailing macro-economic "orthodoxy" is characterized by a clearly discernible historical ebb and flow. This coincides with the so-called long waves of economic history wherein eras of generalized prosperity and rapid growth alternate with periods of slow growth and instability. Predominantly prosperous periods convince economists that a capitalist economy functions quite smoothly provided that the government properly handles its economic responsibilities. The onset of serious economic dislocation, on the other hand, sows the seeds of doubt. The belief that there is a fundamental flaw in orthodox theory spreads and an increasing number of economists become heretics. Interest develops in new theories which offer more plausible explanations for, and resolutions of, the economic crisis.

In the Great Depression John Maynard Keynes challenged the generally accepted view that laissez-faire capitalism provided an adequate economic foundation for society, fighting vigorously against bankrupt policies based on that theory. He also opened up a wide-ranging debate concerning the nature of capitalist macro dynamics and the appropriate form of state intervention in the economy. The economic instability of the 1970's has generated another crisis in economic theory. The prevailing orthodoxy is again under assault, its theory increasingly seen as inadequate, and its policies increasingly viewed as impotent.

The Post-Keynesians have been among the most influential of the anti-orthodox schools. Divided on many questions of theory and policy, they are relatively unified in their belief that what has passed for Keynesian orthodoxy in the United States for a generation is in fact, to use Joan Robinson's phrase, a "bastard" Keynesianism that grossly distorts Keynes' own work. Inspired by Keynes' battle against the status quo in the 1930's, they have been waging what they consider to be the Second War for Intellectual Independence in macro-economic theory and policy.

I. The Two Main "Wings" of Post-Keynesian Economics

The label Post-Keynesian has been applied to a large number of economists who often have little in common beyond an acknowledged intellectual debt to Keynes (although for some Michal Kalecki, David Ricardo, Karl Marx, or Piero Sraffa may be even more influential) and a dissatisfaction with orthodox theory. For descriptive convenience I will use the term "Keynesian" in a very unconventional way, in reference to the widest set of Keynes-influenced critics of neoclassical or bastard Keynesian orthodoxy. Within this set I will distinguish between two major groups, using the term Neo-Keynesians (hereafter N-K) to refer to the largely Cambridge-based or the Cambridge-inspired set of growth theorists such as Robinson, Nicholas Kaldor, and Luigi Pasinetti, who have been building on the seminal work of Keynes, Kalecki, and Roy Harrod, and using the term Post-Keynesians (hereafter P-K) to identify those Keynesians, such as Paul Davidson, Hyman Minsky, G. L. S. Shackle, Sidney Weintraub, and the ubiquitous Robinson, who stress the importance of uncertainty and finance in their work (see Alfred Eichner

*University of Massachusetts-Amherst. I am grateful to Robert Costrell, William Gibson, Douglas Vickers, and especially Leonard Rapping for comments and criticisms.

and Jan Kregel for a more complete score-card).

Keynes never developed the insights of *The General Theory* into an adequate, long-run model. Harrod was the first to incorporate the important conclusions of *The General Theory* in a model of capital accumulation. Writing at the end of the Great Depression, he attempted to prove in a dynamic setting that "sooner or later we will be faced again with stagnation." In his model, the temporary attainment of full employment is possible but not likely, and if reached, is not stable. Moreover, entrepreneurial expectations are a prime determinant of the actual rate of capital accumulation.

The neoclassical growth models first developed by Robert Solow, Trevor Swan, and James Tobin in the prosperous and politically stable 1950's were an attempt to counter Harrod's view of capitalist instability and stagnation. These neoclassical models are dramatically anti-Keynesian in their vision of capitalism. They are balanced, full-employment growth models. Planned savings and planned investment are assumed to be identical *ex ante*—Say's Law holds. Disappointed entrepreneurial expectations are of no consequence. History is irrelevant because capital is instantly and costlessly malleable.

The Neo-Keynesian school has been attacking the realism and logical consistency of the neoclassical theory of capital, growth and distribution ever since. They insist, as did Keynes, that savings does not determine investment, but rather the reverse. The N-Ks take the historic rate of accumulation in any era as exogenously given and ask: What conditions would be required for this rate to be reproduced over time? The class distribution of income plays a key role in the answer. Assuming that propensities to save differ among classes (and institutions), the distribution of income *must* adjust such that the given rate of investment is adequately financed.

Thus, the vision of the capitalist economy reflected in N-K models is fundamentally different from the neoclassical vision. In neoclassical theory the equilibrium rate of growth is determined by the rate of growth of the labor force and the rate of technical change, both exogenous. Factor prices are adjusted via the forces of supply and demand operating in perfectly competitive markets to ensure the continuous full employment of capital and labor. The ratio of capital to labor (or choice of technique) is thus guided by relative factor prices. The equilibrium profit and wage rate are equal to the marginal product of capital and labor, respectively. In N-K theory the exogenous strength of accumulation ultimately determines the equilibrium rate of profit (without reference to marginal productivity) by way of the classical savings function. Thus, N-K theory focuses on the *distributive* role of factor prices in adjusting savings to investment, while neoclassical theory sees them in their *allocative* function as signals of relative factor scarcity to entrepreneurs.

What determines the historic rate of accumulation in N-K theory? Robinson's answer: "To account for accumulation, we have to fall back upon human nature and the structure of society" (p. 15). In Keynesian theory the entire set of economic, political, and social institutions influence the process of accumulation, which itself alters the nature and articulation of these institutions.

The other major wing of the Keynesian revival, the Post-Keynesians, are primarily interested in understanding the process through which investment, savings, and financing decisions are determined in a monetary economy in which the future is uncertain, production takes time, the capital stock is not malleable, and an efficient spot market for durable goods does not exist. This was the economy Keynes analyzed in *The General Theory*. The Hicks-Samuelson counterrevolution in short-run theory replaced it with a counterfeit model that abstracted from these central Keynesian attributes. The P-Ks have helped a new generation of American economists discover the "genuine" *General Theory*.

Keynes believed that in a real world market economy of decentralized decision makers, it is *in principle* impossible for the

economic agents to be provided with the information they require to undertake a set of individual commitments and actions which would ensure continuous full employment. Wealth holders face a crucial decision about the financial forms in which to store wealth over time, and corporations have to make critical judgments concerning real capital investment and its financing. To confidently make optimal decisions in these areas requires complete information concerning the behavior of prices, products, and profits for decades into the future, information which an unplanned economy cannot possibly provide. Concerning these data, Keynes commented "...there is no scientific basis on which to form any calculable probability whatever. We simply do not know" (p. 214). Therefore, the most crucial economic decisions must, by their very nature, be made largely in ignorance and uncertainty. In this uncertain world, it is ephemeral expectations, unpredictable waves of optimism and pessimism, and "animal spirits" which decisively influence investment, and therefore income and employment as well.

These central features of real world capitalism which the P-Ks stress in their analysis make them suspicious of any macro theory characterized by a stable, equilibrium perspective, whether it be neoclassical growth theory, short-run *IS-LM* models, monetarism, or even N-K steady-state growth models such as those of Kaldor or Pasinetti. What emerges most clearly in the P-K vision is instability and disequilibrium.

II. Other Aspects of the Keynesian Critique

It should be clear that there are substantial differences between N-K theory and P-K theory. However, the two wings of the Keynesians agree, though for rather different reasons, that investment demand cannot be represented as a stable, downward-sloping function of the rate of interest as required in both neoclassical growth theory and orthodox short-run macro theory.

The P-Ks have revived the argument presented by Keynes in *The General Theory* to show that the investment function is unstable. Keynes insisted that the marginal efficiency of capital is a function of the expected, future cash flow of investment projects. Because the future is uncertain, entrepreneurial expectations will often be volatile and unstable. Therefore the marginal efficiency of capital is itself unstable and investment demand cannot be considered to be a stable function of the interest rate in real time (see Douglas Vickers).

Neo-Keynesians such as Pierangelo Garegnani, Robinson, and Sraffa have rigorously demonstrated that there can be no logically consistent index of the aggregate quantity of capital associated with a vector of heterogenous capital goods which is independent of the rate of interest. Since any such index will alter in value with every change in the rate of interest, one cannot generate the neoclassical function relating investment demand to the rate of interest through standard supply and demand theory. Capital reversal and the reswitching of techniques of production demonstrate that capital per worker may alternatively rise and fall as the rate of interest declines, proving that investment is not necessarily a monotonic function of the rate of interest.

Robinson has generalized this attack, arguing that the analysis of the process of moving *along* any neoclassical function violates the *ceteris paribus* assumptions logically required for the construction of the curve. A comparison between different long-run equilibrium states is legitimate, but a neoclassical analysis of the process of moving from one to another is not.

If, as the N-Ks argue, investment demand is not a monotonic, inverse function of the rate of interest, then the neoclassical macro-economic vision of supply and demand in competitive markets guiding the allocation of capital and labor through relative factor price movements to ensure balanced full-employment growth is unconvincing. If, as the P-Ks argue, the investment function is unstable, then the Hicks-Samuelson short-run *IS-LM* model is

basically flawed; its comparative static analysis illegitimate. These seemingly unrelated critiques of orthodoxy turn out to be complementary.

There are many other areas of theory in which the Keynesians have made interesting and important contributions. In N-K theory, the process of capital accumulation per se has been placed at the center of economic dynamics, with savings decisions seen as adaptive and largely institutionalized. Unexplained preferences of isolated individuals with respect to consumption patterns over time have been removed from center stage. Capital accumulation is seen as an autonomous, self-sustaining process (see Nina Shapiro). Once it has taken hold in an era, it tends to *create* the micro conditions necessary for its reproduction. Thus, the standard logical relation between microeconomics and macroeconomics is reversed and Marx's question is posed once again: *What is the macro foundation of microeconomics?* Also, the work of those P-Ks who stress the importance of debt and financial intermediation in macroeconomics has been especially helpful in understanding the 1970's, a decade of rising debt and liquidity problems.

In different ways, both the P-Ks and the N-Ks have refocused attention on the issue of income distribution. There is a tendency among P-Ks in particular to explain the inflation of the past decade by reference to distributional struggle among classes. This is a refreshing change from monetarist stories of exogenous increases in the supply of money.

III. Some Problems in Keynesian Theory

For all its accomplishments, Keynesian theory is not without its shortcomings. Foremost among these, in my view, is that taken as a whole, the Keynesian tradition has not made much progress in developing a theory of the complex process through which economic growth, in the short and in the long run, *endogenously* generates those impediments or constraints which prevent its perpetuation. Their theory does not reflect the self-limiting nature of the capitalist growth process nor the functional nature of recessions and depressions in recreating the preconditions for renewed accumulation.

Consider the business cycle. When expansions mature and the "reserve army" of the unemployed declines, a number of impediments to the continuation of growth inevitably develop: the rate of growth of productivity falls below that of real wages causing profit rates and profit margins to decline; debt finance accelerates just when interest rates are rising; the "overinvestment" of the mid-boom leads to excess capacity in many sectors; balance of trade problems develop creating pressure on the exchange rate, etc. The 1960's boom is a textbook example of this process (see the paper by Leonard Rapping and myself). The boom can neither be guided through these obstacles by the market system nor fine tuned around them through policy manipulation. Under existing institutions, as 1974–75 demonstrated once again, recessions remain a necessary condition for the restoration of accumulation.

The idea of an endogenously generated crisis and functional depression has always been at the core of Marxian theory. Keynesians, on the other hand, either assume that accumulation is balanced or stress the potential which exists for instability rather than theorizing about a standard pattern of expansion and contraction. Minsky's excellent work on the relationship of uncertainty and finance to the business cycle (which abstracts from most nonfinancial impediments to growth) is an obvious exception to the general rule.

The typical Keynesian enterprise is an oligopolistic firm which sets its prices by choosing a markup over unit prime cost. Following Kalecki, the markup has traditionally been assumed constant in the short run. Recently, it has even been assumed that firms have the power to vary their markup to assure the adequate financing of planned investment—implying that savings is always equal to planned investment *ex ante*. Any increase in unit labor costs is either passed on, or more than passed on, by the firm in the form of higher prices with no threat to

profits. Thus, financial impediments to accumulation do not exist. This micro model is consistent with N-K growth theory in which attempts to raise the rate of accumulation at full employment generate an accommodating *fall* in the real wage rate and a rise in profit margins, thereby adjusting the amount saved to match the desired increase in investment. Finally, since the Keynesian theory of the firm does not recognize Marx's conceptual distinction between labor and labor power, it has no analytical mechanism through which a more militant work force can cause productivity problems in tight labor markets.

For a substantial subset of the Keynesians then, unless inflation per se is seen as unacceptable, accumulation can roll on forever. The Kaldor-Pasinetti growth models are examples of this genre. Post-Keynesians, of course, emphasize potential instability. Like most Keynesians, however, they fail to see intermittent economic crises as the inevitable outcome of periods of rapid growth as well as a necessary condition for the renewal of the accumulation process.

Post-Keynesian inflation theory is also problematic. Their discussions about distributional struggle are curious because, within the markup pricing framework, distribution is *technically* determined. Powerful labor unions are assumed to be able to push up money wages when labor markets tighten. Corporation managers are seen as too weak to resist labor pressure, so they meekly pass on rising wage costs. Collective labor action is thus assumed to cause inflation and regressively redistribute the wage bill among workers, while leaving the distribution of income between classes unchanged. Traditional anti-inflationary macro policies are considered ineffective or even counterproductive because money wage determination is seen as unaffected by increased unemployment and the markup is unresponsive to, or even rises with, increased excess capacity.

One problem with this Post-Keynesian treatment of inflation is that some of its assumptions are of questionable empirical validity. Profit margins exhibit pronounced cyclical variability. Contrary to Keynesian assertions, relatively full employment typically brings *falling* profit margins and a *rising* wage share of income (see the paper by Raford Boddy and myself). The Keynesian view of "class struggle" is inadequate because it neglects antagonistic class relations within the enterprise. Yet low unemployment reduces labor "discipline" on the shop floor and increases union power to interfere with management's prerogatives with respect to technical change and the organization of the labor process, thereby contributing to the slowdown in the rate of growth of labor productivity that accompanies full employment. The contention that full-employment-induced wage pressures cause an increase in race, sex, and skill income differentials among workers is also factually incorrect. Finally, the picture they draw of the giant multinational corporation as pathetic, impotent, and defenseless in its confrontation with labor is quite unconvincing.

Post-Keynesians also seriously underestimate the impact that deep or prolonged recessions have on inflation. Rising unemployment threatens the job security and alters the working conditions of millions of *employed* workers. It is not only labor's expectations but also their relative power which is affected. Virtually every economic slowdown or recession in the past thirty years has reduced the rate of inflation. The recession of 1974–75 cut the rate of inflation in half. Of course it also threatened domestic and international financial stability, generated widespread protectionism, and led to a political crisis in much of Western Europe. Under current conditions, recession is indeed a dangerous policy tool.

Most Keynesians support some form of incomes policy, but they disagree on the character of the government intervention required. Some support a relatively simple scheme like the tax-based incomes policy. Others propose the complex and ambitious programs suggested by N-K theory, calling for the political determination of the rate and composition of investment, as well as distribution and pricing. They believe that

these programs are achievable through the voluntary association of all major economic interest groups within the political process. In this "pluralist" political process, each group will be willing to participate because it will have at least as much control over the decisions of the proposed town meeting of interest groups as it currently does over developments in the economy (see Eichner).

This political vision is unrealistic. Like Keynes' own theory of politics, it fails to appreciate the immense power that the corporations derive from their exclusive control of capital investment, and thus of jobs and income. If economic conditions do not suit them they stop investing, threatening the economic security of tens of millions of workers. A popular groundswell soon develops in support of whatever government or union concessions are required to induce these corporations to end their "capital strike"—a strike not subject to Taft-Hartley resolution. Why should corporate capital willingly give up the use of this power unless it is given at least as much control over the proposed political institutions which will administratively determine the nature of growth and distribution? The Keynesians could learn a lesson in political realism from Joseph Schumpeter, who argued that investment failed to rebound in the expansion of 1934–37 precisely because corporations conducted a capital strike in protest against what they perceived to be a hostile political, economic, and social environment.

REFERENCES

R. Boddy and J. R. Crotty, "Class Conflict and Macro-Policy: The Political Business Cycle," *Rev. Radical Polit. Econ.*, Spring 1975, *7*, 1–19.

J. R. Crotty and L. A. Rapping, "The 1975 Report of the President's Council of Economic Advisers: A Radical Critique," *Amer. Econ. Rev.*, Dec. 1975, *65*, 791–811.

A.S. Eichner, "A Look Ahead," in his *A Guide to Post-Keynesian Economics*, New York 1979.

______ **and J. A. Kregel,** "An Essay on Post-Keynesian Theory: A New Paradigm in Economics," *J. Econ. Lit.*, Dec. 1975, *13*, 1293–314.

J. M. Keynes, "The General Theory of Employment," *Quart. J. Econ.*, Feb. 1937, *51*, 209–25.

Hyman Minsky, *John Maynard Keynes*, New York 1975.

Joan Robinson, *Essays in the Theory of Economic Growth*, London 1971.

N. Shapiro, "The Revolutionary Character of Post-Keynesian Economics," *J. Econ. Issues*, Sept. 1977, *11*, 541–60.

D. Vickers, "Uncertainty, Choice and the Marginal Efficiencies," *J. Post-Keynesian Econ.*, Winter 1979-80, *2*, 240–54.

DISCUSSION

Peter Kenyon, University of Virginia: My comments on Lorie Tarshis' paper are few. First, Tarshis is to be thanked for his reminder that "normal cost" pricing is quite compatible with profit maximization and standard micro theory. It is timely for Tarshis to remind Post-Keynesians of the importance of the M/A ratio to the functional distribution of income at the micro level—that is, at the level of costs, markups, and prices. (Recall that Sidney Weintraub made exactly the same point over twenty years ago when discussing Kaldor's I/Y theory of profit share.)

Secondly, Post-Keynesians would take issue with Tarshis' comments concerning the investment funds theory of the markup. In Keynesian and Post-Keynesian models, one must be very careful about assumptions as to the state of long-term and short-term expectations and their interaction. As J. A. Kregel argues, Post-Keynesians, in the tradition of Keynes, use a "stationary equilibrium" methodology, where long-period expectations are held constant at a given level, but where short-period expectations may be disappointed and, most importantly, where there is *no* feedback from the disappointment of short-period expectations on the state of long-period expectations.

The investment funds markup model can be seen as an application of this methodology. The state of long-term expectations governs the perception of agents about the evolution of "normal" output, "normal" costs, and the desired stock of capital goods. It is under these conditions that the firm makes its decisions about the desired level of retained profits and about the size of the markup. None of this is inconsistent with maximizing expected long-period profits or with a fairly constant markup or, for that matter, with a variable gearing ratio (although Tarshis' comments concerning sources and uses of funds is well-taken).

In addition, there is no implication that the decision about the desired level of corporate saving has anything to do with corporate investment decisions. In this scheme of things, investment decisions are dominated by the volatile state of short-term expectations, and that is another matter entirely. A careful application of the "stationary equilibrium" methodology would deny any implication that an act of saving would lead to an increase in investment outlays. It is true, however, that Post-Keynesians have been rather careless about this, and Tarshis' concern is well founded.

My third comment concerns two issues raised by Tarshis towards the end of the paper. The first is that Post-Keynesians have paid little attention to the macro-economic implications of micro-adjustment processes. This sin of omission is not restricted to Post-Keynesians; however, Post-Keynesians have been very concerned about the need to analyze processes of adjustment through historical time, and it is incumbent upon them to attempt this analysis. However, this raises severe problems with the Kregel/Keynes methodology discussed above. It is precisely the need to model the feedback from the disappointment of short-period expectations on the state of long-period expectations that is pushed aside by the adoption of the Kregel/Keynes methodology.

The problem is that Post-Keynesians have no model of expectations formation. They constantly appeal to the inoperative concept of "animal spirits." This appeal to capricious fate has led Coddington to dub Post-Keynesian analysis "Chapter 12 Keynesianism." The argument comes from Keynes' early work on probability theory, and says that situations involving decisions about the future are uncertain and cannot be reduced by applying the tools of probability to situations involving risk. Decisions must be made today about a logically unknowable future; however, these decisions will have long-lasting intertemporal consequences.

This is a paradox of Post-Keynesian theory. It wishes to model economic processes that depend upon the formation of, and adjustment to, the state of long-term expectations, but there is no theory of expectations formation. In fact, it is virtually denied that there ever *could* be such a theory: reference is continually made to chapter 12 of

the *General Theory* and to Keynes' *QJE* article. (Incidentally, it would seem difficult to reconcile Kalecki's often-quoted definition of the long run as "...a slowly changing component of a chain of short-period situations..." with the Kregel/Keynes methodology.)

Finally, Tarshis suggests resurrecting Keynes' aggregate supply function. However, without a theory of expectations formation, it is difficult to see how that function could become an operational tool, for it depends critically on a theory of the formation of expectations, a theory absent from the *General Theory*. In any sensible macroeconomic model, expectations must be endogenous, and obviously, there are theories of expectations formation in the literature. It would be to the advantage of Post-Keynesian theory to recognize and adapt these tools if there is to be a viable Keynesian aggregate supply function.

G. C. HARCOURT, University of Adelaide and University of Toronto: I shall concentrate on Janel Yellen's paper, but I shall begin with a few comments on the papers by Lorie Tarshis and James Crotty.

Generally, I found them sympathetic and fair. Tarshis is critical of the micro foundations of Post-Keynesian analysis (hereafter called P-K) because they seem to make business people altruistic, because self-financing is stressed over much, and because they appear to imply certainty, or, at least, more knowledge than business people are likely to have in the world as envisaged by Keynes (and modeled in the *General Theory*). All this sent me back to the paper that Peter Kenyon and I published in 1976 in *Kyklos*, "Pricing and the Investment Decision." If Tarshis will have a look at it, he will find that we are quite close to him on the profit-maximizing point, at least for long-run decisions, that we include a role for external finance, and that we have tried to place our firms in situations of Keynesian uncertainty and hazily formed expectations, which are likely to change and cause a change in behavior as a result.

I agree with both Tarshis and Crotty concerning the major importance of Minsky's work. It provides a model of endogenous cyclical fluctuations which result from the interaction of monetary and real forces in an environment of uncertainty and often unrealised expectations with regard to cash flows. I also agree with Crotty that Marx's insights into the nature of cyclical fluctuations do not get sufficient emphasis in P-K developments, though I think he should have exempted the work of Donald Harris where they are admirably blended with the P-K insights.

Moreover, Kalecki's work on the investment function, which he continuously amended and developed over the whole of his working life, and its extension by Asimakopulos do not leave the theory of investment as open and unfinished as is perhaps implied. I think that Crotty's judgement on the political naivety of consensus policies with regard to the distribution of income and the control of inflation in capitalist economies may well be correct, but what else can a democratic socialist (or even a social democrat) in a country such as Australia, for example, do but try to prove him wrong? I think he is right to stress the over-rigid view in some P-K writings concerning the independence of money wages and money prices from the level of activity, at least over extended periods of time. But I do not think he should push it too far, or he will give undue comfort to the real wage cutters and deflationists generally who have imposed most unnecessary and unjust hardships on wage-earning groups in the advanced capitalist economies for several years now.

Finally, I heartily endorse Crotty's reference to Marx's view that it is macro-economic processes which tend to create required micro-economic structures.

I turn now to Yellen's paper in which the approach of the P-Ks is contrasted with that associated with the standard Keynesian model. First, it is only in the manufacturing sector that the P-Ks argue that "price equals marginal cost [is] the exception rather than the rule." From Kalecki on, Marshallian-type behavior in the agricultural and raw-material-producing sectors has been postulated (wrongly, now, according to Tarshis). Kaldor, in fact, used this distinction for a model of the world economy. It

has Classical roots, in that Smithian-Ricardian increasing returns were associated with the industrial sector, while (historical) diminishing returns were associated with agriculture.

It is true that Eckstein and the others Yellen mentions have provided empirical evidence on normal pricing, but the P-K roots go back much earlier, to Kahn and Kalecki in the 1930's, and to Godley and Neild in the 1950's and 1960's. (Godley and Nordhaus subsequently joined up in the 1970's.)

While I also see much to praise in Eichner's work, his views are not quite as novel as Yellen claims. R. J. Ball, almost completely neglected in the later literature, had already outlined a similar model in 1964, and Adrian Wood's book, *A Theory of Profits*, needs to be mentioned. While "animal spirits" certainly get a good run for their money in the P-K theory of investment, we should not neglect the work of Asimakopulos and Harris, already alluded to above. In any event, "those animal spirits which," as Trevor Swan once put it, "cannot be bottled" are a vital part of the story and reflect an honest imprecision as opposed to what is often only an artificial, rather forced, precision in much orthodox theory.

It is hard to justify either an assumption or a deduction of full employment of the stock of capital goods in a Kaldor-type model. He explicitly assumes that excess capacity is the usual feature of capitalism so that fully employed labor is the bottleneck. With regard to Kalecki's neat aphorism, he did, of course, have chiefly in mind investment expenditure rather than rentier consumption—after all, he was a Marxist: "Accumulate, accumulate... ."

It is hardly surprising that a labor supply function is absent from P-K models, given Keynes's views on the inability of wage earners to determine their real wages (as opposed to the structure of relative money wages). Indeed, given also the Classical *cum* Marxian strands of P-K theory, the notions of schedules and functions are alien to its whole mode of thought—at least, this is true of the Anglo-Italian group who are the more influenced by Sraffa's work. Paul Davidson has taken over much more of Keynes' Marshallian upbringing and has used supply and demand relationships to exposit the lessons of Keynes of the *Treatise* and the *General Theory*—to very good effect.

Yellen's argument for full-employment equilibrium via money wage and price flexibility is also alien to P-K thought. A good exposition is in Paul Wells' recent criticism of Modigliani's presentation of a case similar to Yellen's (see his recent article in the *Journal of Post-Keynesian Economics*). Wells argues that in seeking supposed full-employment-equilibrium price and money wage levels by deflation, the consumption function and investment function would fall in real terms, the liquidity preference function would rise, and there would be a drop in the nominal, and, perhaps, real supply of money in an economy using "inside" money. He concludes that "a serious deflation...cannot but harm the functioning of a money-using, contract, production and exchange economy" (p. 92). A central theme of both Keynes himself and the P-Ks is how essential a stable efficiency wage is for the continued stability of the capitalist system itself.

This leads me to my main disagreement with Yellen's analysis. (I hasten to add that I found her way of putting things and the diagram illuminating and useful.) She argues that P-Ks have been reluctant to do the dynamic analysis because events take place in "historical" rather than in "logical" time. As Kregel explained in his 1976 *Economic Journal* paper, it is the very interdependence of the Keynesian relationships concerning expectations (long- and short-run), when *all* relevant factors are taken into account, that makes the use of *IS-LM* schedules, or the Yellen equivalents, suspect as vehicles for setting out the full Keynesian message. Robert Clower and D. E. Moggridge in their criticisms of the *IS-LM* apparatus come to much the same conclusion.

Thus the P-Ks take a very much "horses for courses" approach, soaking themselves in the actual situations of a given economy at a given point of time in its history, before proceeding with their analysis, or, at least, this is how they *ought* to behave if they have absorbed fully their own central message.

Bias in Economics Education Research from Random and Voluntary Selection into Experimental and Control Groups

By JOHN J. SIEGFRIED AND GEORGE H. SWEENEY*

Evaluations of introductory economics classes employing self-paced instruction, programmed learning, computer-assisted instruction, and other innovative techniques generally report that these pedagogies are no more effective than the traditional methods of instruction in improving college students' understanding of economics (see Siegfried and Rendigs Fels). Students seem to like these innovative teaching methods, however, and consider them to be more effective. Since most of the innovations are based on general principles of learning theory, instructors who adopt them are usually puzzled at the resulting insensitivity of test scores to their innovative efforts.

Various explanations have been offered for the disappointing evaluations of the new teaching methods. First, many of the empirical tests do not control for the level of inputs. If students reduce their effort in courses that adopt new, and purportedly more efficient, teaching methods, the comparison of experimental and control groups will involve a simultaneous change in both pedagogy and input levels. Richard McKenzie and Robert Staaf have argued that a student's response to a change in the relative price of a course (measured by effort required to achieve any given level of success) involves both a substitution effect and an income effect. If introductory courses, which are frequently part of college "distribution requirements," are inferior goods, the income effect would offset the substitution effect. If the income effect were sufficiently large, a successful innovation in introductory economics would consequently be accompanied by a reduction in student inputs, such as study time. Unless the value of the diverted time is properly measured and counted, the evaluation of the experimental teaching method will be inadequate and misleading.

A second explanation for the disappointing results is that measurement of performance has been inadequate (see Judith Yates). Although this argument undoubtedly has merit when only a single dimension of output (cognitive achievement) is used rather than a vector of output components, it does not make sense when most of the innovations are specifically designed to improve cognitive achievement.

A third possible reason for the insensitivity of learning to pedagogical innovation is that the research models employed to test the effectiveness of experimental courses are inadequate. Often the models are *ad hoc*, based more on what data are available to the investigator than on any carefully constructed model of the learning process, and are subject to numerous statistical problems.

Of course, it could be that the innovative teaching methods are truly no more effective than the lecturn and blackboard. This paper demonstrates, however, that the method of selecting experimental and control groups can be the cause of the disappointing evaluations of alternative

*Associate professor and assistant professor of economics, respectively, Vanderbilt University. Ben Bolch, Rendigs Fels, Arthur Goldberger, W. Lee Hansen, Cliff Huang, Allen Kelley, Katherine Maddox, and Mathew Morey provided helpful comments on an earlier draft of this paper. Since we did not take all of their advice, they should all be absolved of any responsibility for remaining errors.

pedagogical techniques even if the new pedagogies are actually superior for certain students.

I. Self-Selection Bias in Economics Education Experiments

Experimental findings depend crucially on the method of assigning students to control and experimental groups. Proper sample selection requires careful consideration of the purpose for offering the experimental section. Specifically, if a department of economics considers offering students the *option* of learning economics by the conventional lecture format or by an innovative pedagogy, an experimental evaluation of this action should *not* be based upon evidence collected from students who were either 1) randomly assigned to experimental and control classes, or 2) allowed to choose among the options themselves.

It is commonly recognized that student learning processes vary across individuals.[1] Students can exploit their comparative learning advantage if numerous pedagogies are offered, enabling each student to choose the teaching technique that he thinks suits his learning process best.

Evaluation of the efficacy of an additional pedagogical option requires measurement of the gain in learning that will accrue to those students for whom the additional option is thought to be most efficient. Therefore, *both the experimental and the control group should consist of only those students who would voluntarily elect the experimental method.* Neither random assignment nor voluntary self-selection of experimental and control groups satisfies this criterion.

To illustrate the nature of the bias resulting from sampling procedure, suppose there exist two different types of students.[2] For simplicity, assume that within each course section, all students of the same type learn the same amount. Let C_r^1 be the cognitive measure of learning of a type 1 student who participates in the section employing regular instructional techniques, and let C_s^1 be the measure of learning of a type 1 student who participates in the section employing the special techniques. Similarly, let C_r^2 and C_s^2 designate the measures of learning of a type 2 student who participates in the regular or special class, respectively.

If the special class is designed to meet the needs of type 2 students, the effectiveness of the special class E might be determined from the difference between the amount which a type 2 student would learn in the special section and the amount which a type 2 student would learn in the regular section:

$$E \equiv C_s^2 - C_r^2 \tag{1}$$

E could be estimated by comparing the mean performance level of students in the special section with the mean performance level of students in the regular section. Let D represent this estimate. If $\bar{C}_r$ and $\bar{C}_s$ designate the mean measure of performance of students in the regular and special class, respectively, then

$$D \equiv \bar{C}_s - \bar{C}_r \tag{2}$$

The estimate D can also be calculated by regressing student performance on a constant and a binary variable which takes a value of one if the student participates in the special section, and zero otherwise. Thus D would equal the coefficient of the dummy variable.

Let the fraction of students in the regular and special class who are of type 2 be represented by f_r and f_s, respectively. The mean

[1]W. Lee Hansen, Allen Kelley, and Burton Weisbrod noted this possibility a decade ago.

[2]Although all students may not judge correctly the most conducive method of instruction, the decision they actually make *given the choice among the alternative pedagogies* is the relevant basis for evaluation since presumably such selection errors will continue if the experimental pedagogy is offered on a permanent basis. In other words, if student's *perceptions* of the relative effectiveness of alternative teaching methods are going to be the permanent basis for allocating them to various sections employing alternative pedagogies, then those perceptions should also be used in the evaluation to identify those who would choose the experimental method.

level of performance of students in each class section is, therefore,

(3a) $$\bar{C}_r=(1-f_r)C_r^1+f_rC_r^2$$

(3b) $$\bar{C}_s=(1-f_s)C_s^1+f_sC_s^2$$

Substituting equation (3) into equation (2), the estimated measure of effectiveness of the special section becomes

(4)
$$D=(1-f_s)C_s^1-(1-f_r)C_r^1+f_sC_s^2-f_rC_r^2$$

Equation (4) can be used to determine the relationship between D and E under various sampling methods.

A. *Random Sampling*

If students are randomly sorted into class sections, the fraction of type 2 students in the special class can be expected to equal the fraction of type 2 students in the regular class. This implies that with random sampling

(5) $$D=(1-f)(C_s^1-C_r^1)+f(C_s^2-C_r^2)$$

or from (1),

(6) $$D=(1-f)(C_s^1-C_r^1)+fE$$

Equation (6) reveals that with random sampling this estimate of E is biased downward if the sample contains any type 1 students (i.e., if $f<1$). Since the special section is not designed for type 1 students, students of type 1 are likely to learn less in that special class. That is, it is likely that $C_s^1 \leqslant C_r^1$. But if the sample contains any students of type 1, $f<1$. Therefore (assuming $E>0$), $D<E$. That is, D understates the true effectiveness of the special class.[3] The smaller the fraction of type 2 students in the sample (i.e., the smaller the value of f), the worse will be the bias.

[3]Assuming type 1 students do better in the regular class (i.e., $C_s^1<C_r^1$), for small enough f, D will be *negative* regardless of the value of E.

B. *Perfect Self-Selection*

Suppose that the special section is carefully advertised and offered as an alternative which students may choose instead of the regular section. If each student recognizes his own type and chooses that section which is better for him, then under perfect self-selection,

(7) $$f_r=0;\ f_s=1$$

Substituting these values for f_r and f_s into equation (4) implies that under perfect self-selection,

(8) $$D=C_s^2-C_r^1$$

Equation (8) reveals that with perfect self-selection the difference between mean performance of the two classes has nothing whatsoever to do with the true measure of effectiveness E! Whether D is positive or negative in this case depends only upon the relative learning abilities of the two types of students in the class best suited to each.

C. *Correct Sampling Procedure*

Equation (4) indicates that D provides an accurate measure of E only if $f_s=f_r=1$. That is, the difference between mean learning levels of the experimental and control groups is an unbiased measure of the effectiveness of a special pedagogy only if both the control and the experimental groups are composed exclusively of students who are the targets of that special pedagogy.

Only if future students will be offered a menu of alternative pedagogical techniques from which they may choose will E be an appropriate measure of the effectiveness of an experimental program. If limited resources prevent the offering of more than one technique, and the experimental course is to be tested as a *replacement* for the conventional course, E is an inappropriate measure.

When the experimental pedagogy is to completely replace the existing teaching method, the effects upon *all* students should be taken into account. A measure of

effectiveness can thus be calculated as a weighted average of the incremental effectiveness of the experimental program for each student type, with weights corresponding to importance of each student type in the population, as defined by equation (5). That is, with random sampling, the variable D measures the average of the disadvantage suffered by type 1 students, weighted by the fraction of the population composed of students of type 1, against the advantage enjoyed by type 2 students, weighted by the fraction of the population composed of students of type 2.

The use of a sample generated by student self-selection is inappropriate in this situation too. The bias caused by employing such a sample can be seen by comparing equations (8) and (5).

Unfortunately, random sampling is difficult to achieve because students control class selection. Although student names may be randomly sorted into class sections, students themselves generally cannot be prevented from switching sections or dropping the course altogether. If a certain type of student is more likely to avoid the experimental section than another, sampling may no longer be random, and the experimental results may be biased.

This problem of bias in the measurement of the relative effectiveness of competing programs results whenever students' participation in one program is correlated with natural ability to learn. Fortunately, there exist methods for correcting this bias when its presence is recognized. The nature of this bias and a brief description of adjustments for it are the subjects of the next section.

II. Selection Bias in Random Samples

Suppose that learning ability varies among students, but the difference between the amount that a student would learn in the experimental course and the amount that he would learn in the regular, or control, course, α, is the same for every student. The amount that the ith student would learn in his economics course is assumed to be represented by the following model, typical of many economics education studies:

$$S_i = W_i + \alpha Z_i \tag{9}$$

where S_i represents a measure of learning by the ith student, W_i represents the ith student's innate ability to learn the subject in the existing or control class, and Z_i is a participation variable which is 1 if the ith student participates in the experimental section, and 0 otherwise.

Let C represent the set of students for which $Z_i = 0$ (the control group), E represent the set of students for which $Z_i = 1$ (the experimental group), n_c equal the number of students in the control group, and n_e equal the number of students in the experimental group. If one estimates α via the regression

$$S = a_0 + a_1 Z \tag{10}$$

the estimated coefficients using ordinary least squares (OLS) regression are

$$\hat{a}_0 = \frac{1}{n_c} \sum_{i \text{ in } C} S_i \tag{11a}$$

$$\hat{a}_1 = \frac{1}{n_e} \sum_{j \text{ in } E} S_j - \frac{1}{n_c} \sum_{i \text{ in } C} S_i \tag{11b}$$

But by equation (9), $S_i = W_i$ for i in C (students in the control group), and $S_j = W_j + \alpha$ for j in E (students in the experimental group). Therefore, (11a) and (11b) may be rewritten as

$$\hat{a}_0 = \frac{1}{n_c} \sum_{i \text{ in } C} W_i \tag{12a}$$

$$\hat{a}_1 = \alpha + \left[\frac{1}{n_e} \sum_{j \text{ in } E} W_j - \frac{1}{n_c} \sum_{i \text{ in } C} W_i \right] \tag{12b}$$

The coefficient $\hat{a}_0$ equals the average learning ability of students in the control group, and the coefficient $\hat{a}_1$ equals the *sum* of the effect of participation in the experimental group (α) *plus* the difference between the average ability of students who participate in the experimental section and the average

ability of students in the control section. *Thus, $\hat{a}_1$ accurately measures the incremental effectiveness of the experimental class only if the average level of ability of students in the experimental group equals the average level of ability of students in the control group.* If a student's ability to learn is correlated with his selection into the experimental section, $\hat{a}_1$ is a *biased* estimator of α. For example, if there is a tendency for better students to avoid the experimental program, $\hat{a}_1$ will underestimate the effect of participation in that program.

The correlation between ability and selection into the experimental class may result from either of two types of phenomena:

(a) both ability and selection are related to some other factor(s) which can be observed and quantified (for example, mathematical ability),

(b) both ability and selection are related to some other factor(s) which cannot be observed or quantified (for example, risk aversion).

Accounting for these correlations requires the assumption of a functional form for the relationships between ability and the other variables, and between selection and the other variables. In practice, linear forms are generally assumed. That is, letting X represent the observable factor, it is assumed that the relationship between innate ability, W, and X may be represented as

$$W_i = b_0 + b_1 X_i + \varepsilon_i \tag{13}$$

where ε_i represents the "error," or relationship between ability and unobserved variables. It is further assumed that the relationship between selection into the experimental section and the value of X may be represented by the equation

$$Z_i = \begin{cases} 1 & \text{if} \quad c_1 X_i + \mu_i \geqslant c_0 \\ 0 & \text{if} \quad c_1 X_i + \mu_i < c_0 \end{cases} \tag{14}$$

where μ_i represents the error or relationship between selection and unobserved variables, and the constants c_1 and c_0 characterize the nature of the relationship between selection and X. If the coefficient c_1 is positive, equation (14) implies that Z tends to equal 1 when X is large and to equal 0 when X is small. Similarly, if the coefficient c_1 is negative, this equation implies that Z tends to equal 1 when X is small and to equal 0 when X is large. (The relative magnitudes of the coefficients c_0 and c_1 specify the precise meaning of the terms large and small.)

If the correlation between ability and selection is due entirely to relationships with observable variables, solution of the problem is easy. (This is the case that ε_i is not correlated with μ_i or X_i.) One need only include the additional variables as independent variables in the learning regression equation. That is, *OLS* estimation of the equation

$$S = a_0 + a_1 Z + a_2 X \tag{15}$$

will yield an a_1 that is an unbiased estimate of α, the incremental effectiveness of the experimental course.[4]

If there is a relationship between ability and an unobserved (or not quantifiable) variable, and a relationship between selection and that same observed variable, *OLS* estimation of equation (15) will yield a biased estimate of the participation effect because selection and ability would still be correlated even after taking into account the relationship with X. (This is the case that ε_i is correlated with μ_i.)

Fortunately, a consistent estimate of the participation effect can be obtained in this more difficult case when the effects of these unobserved variables may be modelled as errors drawn from a bivariate normal distribution. Details of the estimation procedure are provided in Heckman and in Burt Barnow, Glen Cain, and Arthur Goldberger. By assuming that the random variables ε_i and μ_i are drawn from a bivariate normal distribution, one may calculate (up to a factor of proportionality) the influence of

[4]Although this method yields an unbiased estimate of the treatment effect (a_1), the resulting estimate of the sample variance is biased (see James Heckman).

the unobserved correlation as a mathematical function of c_1X and Z. An estimate of the coefficient c_1 may be obtained through estimation of equation (14). (Since Z is a binary variable, a non-linear estimation procedure, such as probit, must be used.) Using this estimate of c_1, the influence of the unobserved correlation may be calculated, and included as an additional independent variable in the regression of S on Z (and X) to obtain a consistent estimate of the treatment effect.

III. Conclusion

When empirically evaluating a pedagogy which is designed *as an additional option* to the conventional lecture and discussion approach, neither random assignment nor student self-selection into regular and special classes is appropriate. When evaluating a competing pedagogy *as a complete substitute* for the conventional method, random selection is the appropriate experimental method. However, biases due to students switching sections and dropping courses are likely to remain.

REFERENCES

B. Barnow, G. Cain, and A. Goldberger, "Issues in the Analysis of Selection Bias," unpublished manuscript, Univ. Wisconsin-Madison, Aug. 1978.

W. L. Hansen, A. Kelley, and B. Weisbrod, "Economic Efficiency and the Distribution of Benefits from College Instruction," *Amer. Econ. Rev. Proc.*, May 1970, *60*, 364–69.

J. Heckman, "Sample Bias as a Specification Error," *Econometrica*, Jan. 1979, *47*, 153–61.

Richard B. McKenzie and Robert J. Staaf, *An Economic Theory of Learning: Student Sovereignty and Academic Freedom*, Blacksburg 1974.

J. J. Siegfried and R. Fels, "Research on Teaching College Economics: A Survey," *J. Econ. Lit.*, Sept. 1979, *17*, 923–69.

J. Yates, "Research in Economics Education: Are Our Horizons Too Narrow?," *J. Econ. Educ.*, Fall 1978, *10*, 12–17.

Pooled Cross-Section, Time-Series Evaluation: Source, Result, and Correction of Serially Correlated Errors

By WILLIAM E. BECKER, JR. AND MATHEW J. MOREY*

Problems associated with serial correlation in statistical estimation and testing of economic cross-sectional relationships have been subjected to time-series analysis by statisticians and econometricians.[1] Such problems, however, have been largely ignored by researchers attempting to evaluate alternative educational treatments via cross-sectional comparisons. This is surprising given that learning takes place over time.[2]

Numerous misspecifications in econometric models involving observations over time are known to bias statistical estimates and jeopardize the validity of statistical testing. While learning is typically measured by educational researchers between two points in time, little or no consideration is given to the manner in which the two observation points are related or to the fact that the two observations are only a "slice" from a sequence of many possible observations over time. Harold Webster and Carl Bereiter may have recognized this when they wrote: "In this derivation, experimental independence of the k measurement trials is assumed in each of the two test administrations. This does not seem to be a hazardous assumption for most mental tests, but it deserves more thorough investigation...we are aware of no studies of this problem..."(p.45).

*Associate and assistant professor of economics respectively, Indiana University-Bloomington. In part this study is based on an earlier study by Becker, supported by a grant from the J. M. Foundation to the Joint Council on Economic Education. We acknowledge suggestions and comments provided by Andrew Ahlegren, W. Lee Hansen, Phillip Saunders, and John Siegfried.

[1]See for example, Charles Henderson and Henri Theil.

[2]Gene Glass et al. provide some examples of an attempt to incorporate time-series techniques in educational research.

This article gives possible sources of the error structures associated with two commonly used cross-sectional education evaluation regression models and to establish the implications of such structures for pre- and posttest evaluation. Contrary to Websters and Bereiter's speculative assumption, the error structures inherent in these two models are shown to cause differing statistical biases depending on the model specification considered and the method of estimation employed. A generalized model is proposed as an appropriate way to handle two particular specification problems and an iterative maximum likelihood estimator is suggested for estimating this model.

I. The Source of Serial Correlation and Its Consequences

As discussed by John Siegfried and Rendigs Fels (p. 929), two regression models often used in economic education program evaluation studies are

(i) the "value-added" or "raw gains" model where adjustment for exogenous covariates in the posttest minus pretest change score is made, and

(ii) the "absolute level" or "regressed gains" model with covariance adjustment of posttest for pretest and for possible exogenous background, for example, socioeconomic variables.

For simplicity and without loss of generality, these two models can be represented as consisting of a single control group and one treatment at a point in time. They may be written respectively as

$$Y_{it} - Y_{it-1} = bX_{it} + \delta D_{it} + e_{it} \tag{1}$$

$$Y_{it} = BY_{it-1} + bX_{it} + \delta D_{it} + e_{it} \tag{2}$$

where all variables are measured in devia-

tion form and

Y_{it} is the posttreatment test score for the ith individual at time t (posttest)

Y_{it-1} is a pretreatment test score for the ith individual at time $t-1$ (pretest)

D_{it} is a dummy variable that is equal to one, if the ith individual was in the experimental or treatment group between pre- and posttesting, and zero if not,

X_{it} is some suspected nonstochastic exogenous source of variation in the ith individual's behavior at time t

e_{it} is the stochastic disturbance term for the ith individual at time t

b, δ, and B are unknown parameters to be estimated over the i individuals, $i=1,2,\ldots,N$.

The parameters in (1) and (2) can be estimated efficiently over i at time t using ordinary least squares (OLS) given the following restrictions on e_{it}, where $E(\ )$ denotes the expectation operator:

$$(3) \qquad E(e_{it})=0 \qquad i \text{ and } t$$

$$(4) \qquad E(e_{it}|X_{it}, D_{it})=0 \qquad i \text{ and } t$$

$$(5) \qquad E(e_{it}e_{js})=0, \qquad i\neq j$$

$$(6) \qquad E(e_{it}e_{js})=0, \qquad s\neq t$$

$$(7) \qquad E(e_{it}e_{js})=\sigma_e^2 \qquad i=j, s=t$$

Conditions (3)–(7) imply that the disturbances have zero mean, a finite constant variance, are uncorrelated across individuals and over time, and are uncorrelated with the exogenous and treatment variables. The assumptions (3)–(7), albeit arbitrary, when true are sufficient conditions to ensure that the OLS estimators of B, b, and δ, are unbiased and efficient.

To establish the manner in which serially correlated disturbances (a violation of condition (6)) may be created in the pre- and posttest evaluation models and to assess the statistical consequences of this for estimation and inference, consider the theoretical relationships underlying (1) and (2). At time t, the ith individual's behavior may be considered a function of the cumulative effect of both the current and a weighted sum of past values of exogenous variables, together with the current and/or some past (experimental) treatments. For example, this behavior may be modeled by a geometrically distributed lag specification,

$$(8) \qquad Y_{it}=b(X_{it}+BX_{it-1}+B^2X_{it-2}+\ldots +B^sX_{it-s}+\ldots) +\delta(D_{it}+BD_{it-1}+B^2D_{it-2} +\ldots B^sD_{it-s}+\ldots)+u_{it}$$

where u_{it} is a random disturbance satisfying the conditions (3)–(7), and B^j, $j=0,1,\ldots,s$, is a coefficient of adjustment which may reflect decay (or growth) in the effect of past values of X on Y over time; $|B|<1$ implies decay ($|B|>1$ implies growth), and $B=0$ implies only a contemporaneous relationship exists between X, D, and Y.

As an example of how (8) might be interpreted, consider a school district that tests students economic understanding when they enter first grade and every spring thereafter. The district uses the outcomes of this testing to measure the effectiveness of alternative first through sixth grade, sequenced curriculum packages. For grades one through six, students use gradewise appropriate materials from only one of two curriculum packages (A or B) for fixed periods of time during the school year. In this case, the treatment variable is a dummy variable such that:

$D_{it}=1$, if student i uses package A in grade t
$=0$, if student i uses package B in grade t

where $t=1,2,\ldots,6$. The variable X_{it} might measure the quantity of time during grade t that the ith student was exposed to one of the curriculum packages. In this model, the ith student's economic understanding at the end of the first grade is only a function of the time spent using a curriculum package in first grade. The ith student's economic understanding in second grade is a function of time spent using a curriculum package in second grade and in first grade. The impact on terminal second grade economic understanding of time spent in first grade will be measured by bB while the effect of second grade time will be measured by just b. If $B<1$ then the time spent in a first grade

curriculum package will have a diminishing effect as the individual progresses through school. (Similarly, it can be seen that terminal third grade understanding will be effected by time spent in first grade on the order of bB^2, and so forth.)

The ith individual's behavior at time $t-1$, or any other lag length, may be represented by

$$(9)\quad Y_{it-1}=b(X_{it-1}+BX_{it-2}+B^2X_{it-3}+\dots)$$
$$+\delta(D_{it-1}+BD_{it-2}+B^2D_{it-3}+\dots)+u_{it-1}$$

In the case where there is only one application of the experimental treatment, $D_{it}=1$ if the ith individual is in the experimental group and zero if in the control group, while $D_{it-j}=0$ for all i and nonzero j; $j=1,\dots,s$.

The only difference between (8) and (9) is the fact that the ith individual has "matured" by one period. The standard evaluation model given in (2) is now derived by simply multiplying (9) by B, subtracting this product from (8), and transposing the BY_{it-1} term,

$$(10)\quad Y_{it}=BY_{it-1}+bX_{it}+\delta D_{it}+e_{it}$$

where the disturbance term, e_{it}, has the structure

$$(11)\quad e_{it}=u_{it}-Bu_{it-1}$$

Using the previous elementary school economic understanding example, it can be seen that equation (10) simply states that terminal t-th grade economic understanding is a function of the previous economic understanding, the amount of time spent using the given curriculum in the t-th grade, the curriculum used, and a first-order moving average disturbance term.

The disturbances in (10) do not possess the simple and desirable characteristics imposed on the e_{it} by conditions (3)–(7). The mean is still zero (i.e., $E(e_{it})=E(u_{it}-Bu_{it-1})=0$) and the variance of the process is constant and is given by

$$(12)\quad \sigma_e^2=E(e_{it}^2)=E((u_{it}-Bu_{it-1})^2)=(1+B^2)\sigma_u^2$$

For the ith individual, however, the covariance between adjacent time periods is no longer zero as required by condition (6); in particular,

$$(13)\quad E(e_{it}e_{it-1})=-B\sigma_u^2\neq 0$$

unless $B=0$ in which case the arguments here are trivial. In what follows, this nonzero covariance will be shown to play a crucial role in estimation and testing of models like (10).

Although only one time period is involved, *OLS* estimation of (2) across individuals will yield biased and inconsistent parameter estimates because condition (6), and thus condition (4), is violated. This is because Y_{it-1} is not independent of e_{it} since

$$(14)\quad E(Y_{it-1}e_{it})=E(e_{it}e_{it-1})$$
$$=-BE(u_{it-1}^2)=-B\sigma_u^2\neq 0$$

Consequently, *OLS* estimators of the e_{it}s and their variance will be biased. Since the *OLS* formulas for the variance of both $\hat{b}$ and $\hat{\delta}$ depend on the estimated variance of the e_{it}s they too will be biased. Under standard distributional assumptions (normality), all test statistics are invalid and should not be employed for inferential purposes or for testing the treatment null hypothesis.

The alternative specification given in (1) is simply a special case of (2) where $B=1$ and Y_{it-1} is not transposed. The distinguishing characteristic in (1) of having Y_{it-1} on the left-hand side, however, is sufficient to remove problems of bias and inconsistency introduced in (10) where BY_{it-1} is on the right-hand side. This is true even though the disturbances in (1) continue to be representative of a moving average process.

The *OLS* cross-individual estimation, at a point in time, of the regression coefficients b and δ will be unbiased in (1) since condition

(4) holds, i.e., $E(e_{it}|X_{iy}, D_{it})=0$. Similarly, e_{it} and its variance can be estimated by *OLS* across individuals, at a point in time, without bias. The *OLS* estimation of the sampling variance of $\hat{b}$ and $\hat{\delta}$, at a point in time, are also unbiased. The treatment null hypothesis may be rejected with statistics having the desirable property of unbiasedness.

II. Assessing Bias in a Multiperiod, Multitreatment Value-Added Model

Reduced-form, single treatment analysis may be meaningful in a value-added model since unbiased coefficients can be estimated using *OLS*. Such is not the case, however, when the experiment is viewed as a multiperiod situation.

For illustrative purposes, consider the simple model in which there are multiple treatment periods but no covariates,

$$(15)\qquad Y_{it}-Y_{it-1}=\delta D_{it}+e_{it}$$

where all variables are defined as in Section I and $e_{it}=u_{it}-u_{it-1}$ with the u_{it}s satisfying conditions (3)–(7). Also let N_1 be the number of individuals receiving the treatment during the experiment (single or multiperiod treatments) where $N_1<N$ and $i=1,2,\ldots,N$. Let T_1 be the number of periods in which treatments were administered to the N_1 individuals where $T_1 \leqslant T$ and $t=0,1,2,\ldots,T$.

If *OLS* estimation on the pooled cross-section, time-series data is employed, the practitioner will inadvertently calculate the variance of $\hat{\delta}$ incorrectly, that is, the variance of $\hat{\delta}$ will be given by

$$(16)\qquad var_{OLS}(\hat{\delta})=\sigma_e^2 \Big/ \sum_1^T \sum_1^N D_{it}^2$$

The bias caused by (16) is evident when comparisons are made with the actual sampling variance of $\hat{\delta}$ which is

$$(17)\quad \sigma^2(\hat{\delta})=\left[\sigma_e^2 \Big/ \sum_1^T \sum_1^N D_{it}^2\right] \times\left[1-\sum_2^T \sum_1^N D_{it}D_{it-1} \Big/ \sum_1^T \sum_1^N D_{it}^2\right]$$

To illustrate the extent of bias that may be created, consider the simplest case where $T_1=T$.[3] Here (17) may be reduced to

$$(18)\quad \sigma^2(\hat{\delta})=\frac{\sigma_e^2}{\bar{D}T(N-N_1)}\left[1-\left(\frac{T-1}{T}\right)\right]$$

When $T_1=T=1$, the term in the brackets is zero; the true sampling variance is the same as the *OLS* formula would suggest. If $T_1=T=2$, *OLS* formulas will overstate the true variance of $\hat{\delta}$ by 50 percent, i.e.,

$$(19)\quad \sigma^2(\hat{\delta})=\frac{\sigma_e^2}{2\bar{D}(N-N_1)}\left(1-\frac{1}{2}\right)$$

$$=.50\; var_{OLS}(\hat{\delta})$$

As T is increased, the upward bias in the *OLS* formula will become worse since the degree of bias is a negative function of T. This overestimation of the variance of $\hat{\delta}$ will cause a like understatement of the t-statistic used to test the null hypothesis that $\delta=0$, $t=\hat{\delta}/\sigma(\hat{\delta})$. Thus, the inappropriate use of the *OLS* formulas will increase the likelihood of committing a Type II error, that is, failure to reject a false null hypothesis.[4]

III. Controlling for Serial Correlation in a Quasi-Value-Added Model

Given the shortcomings of the value-added and absolute level models, but recognizing their usefulness in evaluation, a third type of "quasi-value-added" model seems warranted. Such a model would be of the

[3]A more detailed account of the extent to which the *OLS* variance formula will be biased is contained in an expanded version of this paper, and is available from the authors upon request.

[4]One may also wish to note that a value-added model violates the "invertibility condition" for moving average processes since the u_{it-1} coefficient is not between -1 and $+1$.

form

$$(20)\quad \Delta Y=\begin{bmatrix}\Delta Y_{11}\\ \Delta Y_{12}\\ \vdots\\ \Delta Y_{1T}\\ \vdots\\ \Delta Y_{NT}\end{bmatrix}=\beta\begin{bmatrix}X_{11}\\ X_{12}\\ \vdots\\ X_{1T}\\ \vdots\\ X_{NT}\end{bmatrix}+\delta\begin{bmatrix}D_{11}\\ D_{12}\\ \vdots\\ D_{1T}\\ \vdots\\ D_{NT}\end{bmatrix}+\begin{bmatrix}e_{11}\\ e_{12}\\ \vdots\\ e_{1T}\\ \vdots\\ e_{NT}\end{bmatrix}=Z\gamma+e$$

where

$$\Delta Y_{it}=Y_{it}-BY_{it-1},\qquad Z=[X,D],$$

$$\gamma=\begin{bmatrix}\beta\\ \delta\end{bmatrix},\qquad \text{and}\quad e_{it}=u_{it}-Bu_{it-1}$$

For any value of B between zero and one, a generalized least squares estimator of γ is now given by

$$(21)\qquad \hat{\gamma}=(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta Y$$

where Ω is an $NT\times NT$ matrix defined by

(22)

$$\Omega=\left[\begin{array}{ccccccc|cccc}
1+B^2 & -B & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
-B & 1+B^2 & -B & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
0 & -B & 1+B^2 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0\\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \\
0 & 0 & 0 & \cdots & 1+B^2 & -B & 0 & \cdots & 0 & 0 & 0\\
0 & 0 & 0 & & -B & 1+B^2 & -B & \cdots & 0 & 0 & 0\\
0 & 0 & 0 & & 0 & -B & 1+B^2 & \cdots & 0 & 0 & 0\\
\hline
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots\\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1+B^2 & -B & 0\\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -B & 1+B^2 & -B\\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & -B & 1+B^2
\end{array}\right]$$

Since the "best" value of B for use in Ω is not known a priori, a search procedure must be initiated to find the value of B that minimizes the weighted sum of squared residuals $\hat{e}'\,\Omega^{-1}\,\hat{e}$. The estimator of γ obtained for the "best" value of B will be unbiased with an unbiased dispersion matrix given by

(23)

$$var(\hat{\gamma})=\hat{e}'\Omega^{-1}\hat{e}(Z'\Omega^{-1}Z)^{-1}/(TN-2)$$

Estimation of the relevant coefficients in (20) by (21), and their variance by (23), does capture the essence of the absolute level model since the coefficient of Y_{it-1} is not equal to unity. Unlike the absolute level model, Y_{it-1} remains on the left-hand side of (20) where its coefficient is estimated iteratively between zero and one. Thus, we do not have to worry about getting a biased estimate of B, as was the case in an absolute level model. Finally, for as few as two periods, the value-added model was shown to yield 50 percent overestimation of the variance of the treatment coefficient. The quasi-value-added model will yield unbiased variance estimates regardless of the number of periods employed. Estimation of coefficients in (20) by (21), and their variances by (23), is computationally cumbersome, but the subsequent elimination of bias should result in more credible inferences.

REFERENCES

Gene Glass et al., *Design and Analysis of Time-Series Experiments*, Boulder 1975.

C. Henderson, "Comment on 'The Use of Error Components Models in Combining Cross Section with Time Series Data'," *Econometrica*, Mar. 1971, *39*, 397–401.

J. Siegfried and R. Fels, "Research on Teaching College Economics: A Survey," *J. Econ. Lit.*, Sept. 1979, *17*, 923–69.

Henri Theil, *Principles of Econometrics*, New York 1971.

H. Webster and C. Bereiter, "The Reliability of Changes Measured by Mental Test Sources," in Chester Harris, ed., *Problems in Measuring Change*, Madison 1967, p.45.

Guessing and the Error Structure of Learning Models

By MICHAEL K. SALEMI AND GEORGE E. TAUCHEN*

This paper is broadly concerned with problems associated with the use of test score data to infer the relative strength of inputs to the production of learning. It should be of interest to those employing a typical strategy of economic education research: pretesting students, applying some special educational treatment to a subset of students, and posttesting the students. By this strategy the researcher enquires whether the treated group learned more or learned more efficiently. This paper specifically addresses several concerns raised by Thomas Johnson by laying out a set of structural equations and thereby modelling the probability of answering a test item correctly and by dealing in a novel way with the hypothesis that some test questions are more difficult than others.

The key structural equation of the model is the learning production function which explains student learning in terms of student aptitude and study time. Because student aptitude cannot be observed, a standard procedure has been to use the pretreatment test score as an aptitude proxy. William Becker and Salemi have pointed out that such a procedure can seriously bias estimates of the production function parameters. The approach of this paper is to model explicitly the probability distribution of the test score of a student conditional on his ability level. This approach is similar to that used to estimate models with qualitative dependent variables. But the test score model suggests a convenient approximation and a correction to least squares which together deliver consistent estimates of the structural parameters.

I. The Structural Equations

A. *The Test Score Model*

It is assumed that a student's ability to master economic concepts can be represented by a one-dimensional index, A. Although A cannot be observed directly, precourse test data permit the researcher to control for A. Consider a multiple choice test containing M questions divided on a priori grounds into K difficulty categories with M_k questions in the kth category.[1] Following Frederick Lord and Melvin Novick, define the indicator variable I and α_k so that

$$(1a)\qquad I\begin{cases} \geqslant \alpha_k & \text{if a student knows the answer to a question of difficulty level } k \\ <\alpha_k & \text{otherwise} \end{cases}$$

Further define the dichotomous variable Q_{jk} so that

$$(1b)\qquad Q_{jk}\begin{cases} =1 & \text{if the student answers the } j\text{th question in the } k\text{th difficulty category correctly} \\ =0 & \text{if not} \end{cases}$$

*Assistant professors, University of North Carolina-Chapel Hill and Duke University, respectively. This paper reports the highlights of work discussed in more detail in a paper available from us on request. We thank but absolve from responsibility Bill Becker, Lee Hansen, Tom Johnson, John Siegfried, and the members of the North Carolina State University econometrics seminar. We thank Augusto Lopez-Claros for competent help in carrying out the estimation.

[1]We have the Test of Understanding of College Economics (*TUCE*) in mind. The *TUCE* designers identify three difficulty levels of questions: those requiring students to recognize and understand economic concepts; those requiring students to apply a single concept; and those requiring students to apply concepts in a complex way.

The distinction between I and Q_{jk} is important because a student will generally guess the answer when he does not know it. The link between aptitude and the indicator function is given by

$$(2) \qquad I = A + U$$

where U is a random variable with symmetric cumulative distribution function F.[2] Equation (2) says that whether or not a student knows the answer to a test question depends both on the student's aptitude and a variety of omitted factors the total effects of which are represented by U.

Define $\pi_k(A)$ to be the probability that a student with aptitude A correctly answers a question of difficulty level k. Let λ be the probability that he guesses correctly if he does not know the answer. It is assumed that λ does not vary with aptitude or with the difficulty level of the question. Using the symmetry of F one may deduce

(3)

$$\pi_k(A) \equiv Prob\ (Q_{jk} = 1) = \lambda + (1 - \lambda)F(A - \alpha_k)$$

which says that the probability that a student answers correctly a question with difficulty level k is positively related to A and λ, and inversely related to α_k.

Define S_k to be a student's score on the questions comprising the kth difficulty category. We assume that S_k is the sum of M_k independent Bernoulli trials each with probability of success equal to $\pi_k(A)$. Then, conditional on A, S_k is approximately normal with mean of $M_k\pi_k(A)$ and variance of $M_k\pi_k(A)(1 - \pi_k(A))$. As seen in (3), S_k is a non-linear function of aptitude. To put the model in linear form, Taylor approximate F about zero, the mean value for U.

$$(4) \quad F(A - \alpha_k) \doteq F(0) + F'(0)(A - \alpha_k)$$

[2]Specifying properties of the cumulative distribution function F amounts to choosing units in which to measure A. Below we impose $F(0) = .50$ and $F'(0) = (\sqrt{2\pi})^{-1}$ which imply that the difference between A and α_k is measured in units of standard deviations of U.

Equations (3) and (4) and the normality of S_k imply that the relationship between aptitude and test score is

$$(5) \qquad S_k = C_{1k} + C_{2k}A + e_k$$

where $C_{1k} \equiv M_k(\lambda + (1 - \lambda)F(0) - (1 - \lambda)F'(0)\alpha_k)$, $C_{2k} \equiv M_k(1 - \lambda)F'(0)$, and where $E(e_k|A) = 0$. It is important to our approach that (5) holds for both pretest and posttest scores.

B. *The Learning Production Function*

It is assumed that the learning process may be thought of as a production process in which the output is increased economics aptitude. The inputs to the learning process are student time, student ability at the beginning of the course, and a vector of characteristics describing the learning environment.

We suppose that the production function has the form

$$(6) \quad L \equiv A' - A = \theta_1 A + \theta_2 \mathrm{T} + \theta_3 EN + \theta_4 A \cdot \mathrm{T} + \theta_5 A \cdot EN + \theta_6 \mathrm{T} \cdot EN$$

where learning L is the increment to precourse aptitude A; T is student study time, EN is a vector of variables describing the learning environment, and the θ_j are the parameters of the learning production function (θ_3, θ_5, and θ_6 are vectors conformable with EN). A priori one would expect to find θ_1, θ_2, and θ_4 greater than zero. Of particular interest are the marginal learning product of aptitude equal to $\theta_1 + \theta_4\mathrm{T} + \theta_5 \cdot EN$ and the marginal learning product of student study time equal to $\theta_2 + \theta_4 A + \theta_6 EN$. Each of these terms has components representing the direct marginal contribution of the factor, the marginal contribution due to the interaction of aptitude and time, and terms which permit the marginal product to be different in various learning environments.

Note that (6) assumes an exact production function so that all the error in the model is due to U in equation (2). This assumption simplifies estimation of the model since it implies that the error in the

test score equation, (5), has the same distribution in the postscore and prescore cases.

II. The Corrected Least Squares Procedure

The basic idea of this section is that the use of student precourse test scores as a proxy measure for aptitude requires a correction to least squares in order to achieve consistent estimation of the production function parameters. The correction is based on certain population regressions implied by the model and amounts to a pretreatment of the data. In the psychometric literature, population regressions are often used to address the question of correcting test scores to obtain more accurate rankings of students by ability. In contrast, we use these population regression corrections not to rank students but rather to obtain consistent estimates of the parameters of interest. A further point developed in this section is that a priori information which permits test questions to be grouped into difficulty categories also permits a test of the model's validity because it introduces restrictions across the parameters of the regression equations.

A. *The Prescore as a Proxy for Aptitude*

Equation (5) may be inverted to yield aptitude as a function of a student's pretest score.

$$A=(S_k-C_{1k}-e_k)/C_{2k} \tag{7}$$

There are two points to be made by way of interpreting (7). First, S_k is an error-ridden proxy for aptitude as the presence of e_k in (7) makes clear. Second, explicitly taking account of guessing via the parameter λ alters the implication for aptitude of a change in S_k. For example, the model predicts that a one point increase in the pretest score *ceteris paribus* should be interpreted as an aptitude increase of $(M_k \cdot F'(0)(1-\lambda))^{-1}$.

To derive the relationship between the posttreatment test score and the pretreatment test score, substitute (7) and (6) into the version of (5) valid for the posttreatment score and obtain

$$\begin{aligned} S'_k = & -C_{1k}\theta_1+(1+\theta_1)S_k \\ & +(C_{2k}\theta_2-C_{1k}\theta_4)\mathrm{T}+(C_{2k}\theta_3-C_{1k}\theta_5)EN \\ & +\theta_4\mathrm{T}\cdot S_k+\theta_5 EN\cdot S_k+C_{2k}\theta_6\mathrm{T}\cdot EN \\ & +e'_k-(1+\theta_1+\theta_4\mathrm{T}+\theta_5 EN)e_k \end{aligned} \tag{8}$$

for $k=1, 2,\ldots,K$. Equation (8) is the model's prediction concerning the interrelationship among the observable variables but it is not a proper regression equation since by (5) e_k and S_k are correlated. In (5), $E(e_k|A)=0$, but in (8) the conditional expectation of e_k given S_k, T, and EN is not, and it is this fact which implies that estimation of (8) in its uncorrected form will not yield consistent estimates of the parameters of interest.

B. *The Correction to Least Squares*

It is possible to derive an expression for $E(e_k|S_k,\mathrm{T},EN)$ and thus to write a corrected version of equation (8). Assume that T and EN are strictly exogenous variables and are thus uncorrelated with e_k. Next write

$$E(e_k|S_k,\mathrm{T},EN)=b_{0k}+b_{1k}S_k+b_{2k}EN+b_{3k}\mathrm{T} \tag{9}$$

Equation (9) is the population regression of the error from the pretest score equation on the observed explanatory variables. Our strategy is to employ (9) to correct (8). For the general errors in variables problem this strategy will fail since $Cov(e_k,S_k)$ would be unknown; however, the structure of the test-score model implies

$$Cov(S_k,e_k)=\left(M_k^2\bar{\pi}_k-M_k^2\bar{\pi}_k^2-\sigma_{S_k}^2\right)/(M_k-1) \tag{10}$$

where $\bar{\pi}_k$ is a sample mean and $\sigma_{S_k}^2$ a sample variance. Equation (10) together with the sample covariances of S_k, EN, and T are sufficient information to estimate the b_{jk}.

Use (9) together with (8) to obtain

$$(11)\quad E(S_k'|S_k,\mathrm{T},EN) = -C_{1k}\theta_1 + (1+\theta_1)\tilde{S}_k + (C_{2k}\theta_2 - C_{1k}\theta_4)\mathrm{T} + (C_{2k}\theta_3 - C_{1k}\theta_5)EN + \theta_4\mathrm{T}\cdot\tilde{S}_k + \theta_5 EN\cdot\tilde{S}_k + \theta_6 C_{2k}\mathrm{T}\cdot EN$$

where $\tilde{S}_k = S_k - b_{0k} - b_{1k}S_k - b_{2k}EN - b_{3k}\mathrm{T}$. A comparison of equations (11) and (8) makes it clear that a correction to the least squares procedure is required to produce consistent estimates of the structural parameters. For example, (11) shows that the probability limit of the uncorrected least squares coefficient of S_k is $(1+\theta_1)(1-b_{1k})$ rather than $(1+\theta_1)$ as (8) might seem to suggest.

C. *Parameter Identification and the Estimation Procedure*

To simplify notation rewrite equation (11) as

$$(12)\quad E(S_k'|S_k,\mathrm{T},EN) = B_{k0} + B_{k1}\tilde{S}_k + B_{k2}\mathrm{T} + B_{k3}\cdot EN + B_{k4}\mathrm{T}\cdot\tilde{S}_k + B_{k5}EN\cdot\tilde{S}_k + B_{k6}EN\cdot\mathrm{T}$$

For $K=3$, (12) is a system of three equations and twenty-one coefficients. However, the twenty-one coefficients are each functions only of nine structural parameters ($\theta_1, \theta_2, \ldots, \theta_6, \alpha_1, \alpha_2, \alpha_3$) and the three normalizing parameters (λ, $F(0)$, $F'(0)$). The restrictions that the B_{kj} obey are apparent from a comparison of (11) and (12). Estimation of (12) with and without the restrictions in force provides a natural test of the validity of the model.

The procedure we suggest to estimate the parameters is comprised of three parts. First estimate the b_{kj} of (9) and form the $\tilde{S}_k$. Recall that this strategy is available to us precisely because our test score model tells us how e_k and S_k covary. Second, estimate (12) by ordinary least squares using the corrected prescores but not imposing the restrictions. Estimates of the residuals from these regressions can be used to test for heteroskedasticity and, if needed, to obtain consistent estimates of the variances of the error terms. The sums of squared residuals of these unconstrained regressions are also needed to form a test of the validity of the restrictions. Third, estimate the restricted model by an appropriate non-linear procedure.

III. Empirical Results

The results of applying our suggested estimation procedure to the data set described in Becker and Salemi are briefly summarized in Table 1. Because preliminary estimation suggested that the interactions between study time and environment and between study time and aptitude provided for in (6) were unimportant, they were excluded from the regressions reported. Two models were estimated. In Model I, the production function includes ability, study time, and dummy variables designating the school attended by the students. In Model II, the production function includes the Model I variables plus interactions between the school dummies and aptitude. Unconstrained estimates with both aggregated ($M=33$) test score data and disaggregated data ($M_k=11$, $k=1$, 2, 3) were obtained by ordinary least squares. Constrained estimates with the disaggregated data were obtained by full-information maximum likelihood (*FIML*) using a version of the Davidon-Fletcher-Powell algorithm described by Karl Jöreskog. The χ^2 test for heteroskedasticity suggested by T.S. Breusch and A. R. Pagan was performed using the corrected prescore data in both aggregated and disaggregated cases. These tests fail to reject the null hypothesis of no heteroskedasticity at standard significance levels.[3]

[3]Inspection of (8) will suggest immediately that heteroskedasticity may be a problem in this model. However, as we show in our more detailed paper, heteroskedasticity is likely to be less serious a problem when the errors in variables problem is more serious. For our data the errors in variables problem is quite serious and thus we are not surprised by the results of these tests.

TABLE 1—ESTIMATES FOR THE COEFFICIENTS AND STANDARD ERRORS OF ABILITY (θ_1) AND STUDY TIME (θ_2) IN THE LEARNING PRODUCTION FUNCTION

Model: Estimation Procedure		$\hat{\theta}_1$	$\hat{\theta}_2$
	Becker and Salemi	.22(.23)	.05(.14)
A.	Aggregated Test Scores Unconstrained *OLS*		
	I:Uncorrected	−.307(.065)	−.003(.030)
	I:Corrected	.488(1.61)	−.008(.099)
	II:Uncorrected	−.307(.212)	−.009(.030)
	II:Corrected	.502(.524)	−.014(.030)
B.	Disaggregated Test Scores Unconstrained *OLS*		
	I:Corrected ($k=1$)	12.82(2.21)	.048(.015)
	I:Corrected ($k=2$)	2.63(.914)	−.034(.017)
	I:Corrected ($k=3$)	.326(.271)	.006(.015)
	II:Corrected ($k=1$)	9.44(5.36)	.050(.016)
	II:Corrected ($k=2$)	2.72(2.46)	−.035(.017)
	II:Corrected ($k=3$)	.386(.825)	.006(.015)
C.	Disaggregated Test Scores Constrained *FIML*		
	I:Corrected	.368(.217)	−.003(.008)
	II:Corrected	.504(.300)	−.003(.008)

Notes: $NOBS=314$. The *FIML* computations impose the normalizations $\lambda=.25$, $F(0)=.50$, and $F'(0)=(\sqrt{2\pi})^{-1}$. Becker and Salemi results were computed from their Table 3, p. 84, and were obtained by them via an instrumental variables approach.

The four major findings of our econometric work are as follows: First, when uncorrected test score data are employed the *OLS* estimates of θ_1 are negative and significant. However, with the corrected data the estimates of θ_1 are positive as economic theory would predict and marginally significant. Positive estimates of θ_1 are obtained whether Model I or Model II is estimated, whether aggregated or disaggregated data are employed, and whether or not the constraints are imposed. In every case the estimated value of θ_1 is greater than that reported by Becker and Salemi. For Model I, θ_1 is the marginal product of aptitude. For Model II, however, the marginal product of aptitude is permitted to vary across schools and is equal for school j to $\theta_1+\theta_{5j}$. Using *FIML* on the disaggregated data it was found that the marginal product estimates ranged between .680 (.376) and .206 (.163).

Second, when the disaggregated data are employed imposing the restrictions implied by the model disciplines the estimation in an important way as a comparison of parts B and C in Table 1 readily makes clear. On the one hand then, the restrictions have content important for extracting the structural parameters. On the other hand, the data do not seem to accord well with these restrictions. A comparison of the sum of squared residuals in the constrained and unconstrained cases was performed using the standard likelihood ratio test. The appropriate statistic is estimated to be 67.53 which just exceeds the ninety-ninth percentile of the χ^2 distribution with 42 degrees of freedom.

Third, the model returns estimates of the *alphas* from equation (1a) exactly in accordance with the view that category 1 questions require the least aptitude to master and category 3 questions the most. The actual levels for the *alphas* (which depend upon the normalizations reported in Table 1) are for Model II: 2.58(1.28), 2.74(1.22), and 3.60(1.43), respectively. An examination of a joint confidence ellipse suggests that the difference between *alpha* 1 and *alpha* 2 is statistically small, while that between *alpha* 2 and *alpha* 3 is large.

Fourth, somewhat surprisingly our estimates of θ_2, the marginal product of time in the learning production function, are insignificantly different from zero except in the unconstrained disaggregated case. This may be due in part to the fact that students reported their own study time.

In conclusion, we would stress that the prescore correction suggested here provides

a useful alternative to instrumental variables procedures for estimating learning models. We also take the view that our results recommend further studies using disaggregated test score data to estimate parsimonious structural models. We conjecture that a useful alternative to our hypothesis of a constant guessing parameter, λ, is to model λ as a function of student aptitude. Finally, we recommend that further research be designed to measure student study time more directly.

REFERENCES

W. Becker and M. Salemi, "The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models," *J. Econ. Educ.*, Spring 1977, *8*, 77–92.

T. S. Breusch and A. R. Pagan, "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, Sept. 1979, *47*, 1287–94.

T. Johnson, "Research on Economic Education: How Well is it Answering the Questions Asked?," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 22–27.

K. Jöreskog, "A Computer Program for Minimizing a Function of Several Variables," mimeo., Educational Testing Service, Princeton 1973.

Frederic Lord and Melvin Novick, *Statistical Theories of Mental Test Scores*, Reading, Mass. 1968.

DISCUSSION

GARY CHAMBERLAIN, University of Wisconsin-Madison: The paper by John Siegfried and George Sweeney provides some useful insights into the selection biases that can arise in evaluating alternative teaching techniques. Say that there is a regular program ($x=0$) and a special program ($x=1$). Each individual is characterized by a performance level under the regular program (y_0) and under the special program (y_1). Initially we observe a mean score under the regular program. This is the sample counterpart to $E(y_0)$. Then a special program is introduced and students are allowed to choose between the two techniques, providing estimates of $E(y_1|x=1)$ and $E(y_0|x=0)$. Siegfried and Sweeney point out that $E(y_1|x=1)-E(y_0|x=0)$ in general differs from $E(y_1)-E(y_0)$ if students choose the program that is more effective for them. A more novel point is that $E(y_1) - E(y_0)$ is not necessarily a relevant measure of the program's effectiveness. It is for a program that will replace the existing one, but not if both programs are to be offered, with students free to choose between them. In that case a relevant measure is $E(y_1|x=1)-E(y_0|x=1)$, the mean increase in performance of those who elect the special program. Siegfried and Sweeney note that in general this also differs from the commonly used $E(y_1|x=1)-E(y_0|x=0)$.

I would like to point out that $E(y_1|x=1) -E(y_0|x=1)$ is in fact readily estimable. We have a direct estimate of $E(y_1|x=1)$ and only need an estimate of $E(y_0|x=1)$. Note that $E(y_0)=pE(y_0|x=1)+(1-p)\cdot E(y_0|x=0)$, where p is the probability of selecting the special program. Since p, $E(y_0)$, and $E(y_0|x=0)$ have sample counterparts, we can obtain an estimate that almost surely converges to $E(y_0|x=1)$ as the sample size increases.

Siegfried and Sweeney also consider estimation of $E(y_1)-E(y_0)$ when random assignment is not feasible. If the probability of using the new technique depends only on y_0 and not on the gain (y_1-y_0), then our previous estimate of $E(y_1|x=1)-E(y_0|x=1)$ is still appropriate. The more difficult case is the quite plausible one in which the choice of technique depends on y_1-y_0. As Siegfreid and Sweeney note, a novel technology exists for dealing with this case, at least under normality assumptions. I was disappointed that no applications of this approach were presented in the papers. It should be stressed, however, that the necessary distributional assumptions are not innocent. The model is not identified without specific distributional assumptions.

To see this, consider the simple case in which $x=1$ if $y_1 \geqslant y_0$ and $x=0$ if $y_1<y_0$. We observe only x and $max(y_0,y_1)$. The observed data is not affected if y_0 is reduced by an arbitrary amount when $x=1$, or if y_1 is reduced by an arbitrary amount when $x=0$. The empirical distribution of $x, max(y_0,y_1)$ is not affected, and passing to the population we see that a given distribution for x, $max(y_0,y_1)$ is consistent with any value for $E(y_1)-E(y_0)$.

One might hope that observing an instrumental variable (z) that is independent of y_1-y_0 but correlated with x would help. Say that $x=1$ if $y_1-y_0-z\geqslant 0$, so that there is observable variation in the threshold. Once again we can decrease y_0 by an arbitrary amount if $x=1$, and we can decrease y_1 by an arbitrary amount if $x=0$. The population distribution for x,y_x,z is consistent with any value for $E(y_1)-E(y_0)$; there is still no identification without specific distributional assumptions.

So if we proceed in this direction, some sensitivity analysis is called for. This can lead to a firm negative conclusion if, for example, going from the normal to the student family makes a big difference. The more difficult question is how to settle on a class of distributions that is large enough.

William Becker and Mathew Morey consider some of the measurement issues that arise when test scores are available at several points in time. They are particularly concerned with the bias in regressions of scores on previous scores. In their framework, the previous score enters through a Koyck transformation of a geometric distributed lag model, with the same lag struc-

ture imposed on all of the variables. They assume no serial correlation in the original model (hence no individual specific variance components), and so there is negative serial correlation in the transformed model. The bias problem is potentially serious, but the proposed solution is consistent only as the number of time points (T) increases. To see this, say that $T=2$; then their estimator reduces to a least squares regression of the second score on the first score and other variables. This is the biased estimator that is being criticized. The problem is that the initial score distribution is omitted in specifying the likelihood function.

Typically there will only be a few measurements over time, so that much of the information is coming from the independent replication across individuals. In making asymptotic arguments, we should consider limiting distributions in which the number of individuals increases but not the amount of information on an individual. Restrictions that are standard in analyzing a single long time-series can be relaxed. We can drop stationarity assumptions and allow for individual specific parameters (dispensing with ergodicity).

I would like to consider an appropriately general starting point. For each individual, there is a set of scores taken at different points in time: $\mathbf{y}'=(y_1,\ldots,y_T)$, and a set of variables indicating the type of instruction: $\mathbf{x}'=(x_1,\ldots,x_T)$. We shall simplify by suppressing other measured inputs, such as the amount of student time. We can consider the conditional mean of the scores given all of the technique variables (or more generally, the conditional distribution): $E(\mathbf{y}|\mathbf{x})=\mathbf{\Pi x}$, where $\mathbf{\Pi}$ is a $T\times T$ matrix of regression coefficients.

How shall we interpret future values of $\mathbf{x}$ having nonzero coefficients in $\mathbf{\Pi}$? One possibility is that there are unmeasured characteristics (α) of the individuals that are correlated with choice of instructional technique. Say that $E(y_t|\mathbf{x},\alpha)=\Sigma_j\beta_{tj}x_{t-j}+\alpha, E(\alpha|\mathbf{x})=\boldsymbol{\delta}'\mathbf{x}$, so that $\mathbf{\Pi}=\mathbf{B}+\mathbf{1}\boldsymbol{\delta}'$, where $\mathbf{B}$ is a lower triangular matrix representing an unconstrained distributed lag and $\mathbf{1}$ is a $T\times 1$ vector of ones. (The problem of missing initial values for $\mathbf{x}$ is nontrivial but will not be dealt with here.)

Consider the transformation from $\mathbf{y}$ to $(y_1,y_2-y_1,\ldots,y_T-y_1)=(y_1,\Delta\mathbf{y}')$. Then we have $E(\Delta y_t|\mathbf{x})=(\beta_{t,t-1}-\beta_{10})x_1+\beta_{t,t-2}x_2+\ldots+\beta_{t0}x_t$; all of the distributed lag coefficients are identified except for the level of the x_1 coefficients. If we specify that the residual covariance $V(\mathbf{y}|\mathbf{x})$ is not a function of $\mathbf{x}$, then we can allow it to be a free $T\times T$ covariance matrix ($\mathbf{\Sigma}$). A multivariate regression analysis of $\Delta\mathbf{y}$ is straightforward, and we can test whether $E(\Delta\mathbf{y}|\mathbf{x})$ gives a lower triangular matrix of regression coefficients. Under normality assumptions the maximum likelihood estimator of $\mathbf{B}$ based on the distribution of $\Delta\mathbf{y}$ (given $\mathbf{x}$) is in fact identical to the maximum likelihood estimator based on $\mathbf{y}$; for the conditional distribution of y_1 given $\Delta\mathbf{y}$ places no restrictions on the parameters of the $\Delta\mathbf{y}$ distribution if $\boldsymbol{\delta}$ and $\mathbf{\Sigma}$ are unconstrained. The transformation to $\Delta\mathbf{y}$ is useful since it shows that we do not have to assume linearity for the auxiliary regression $E(\alpha|\mathbf{x})$.

The model can be extended to allow the effect of α to vary over time. We can also allow for different types of "ability" with the relative effects changing over time: $\mathbf{\Pi}=\mathbf{B}+\boldsymbol{\lambda}_1\boldsymbol{\delta}_1'+\boldsymbol{\lambda}_2\boldsymbol{\delta}_2'$, where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are free $T\times 1$ vectors. Once again we would like to avoid assuming that the auxiliary regression functions $E(\alpha|\mathbf{x})$ are linear; rather we would like to interpret $E(\alpha|\mathbf{x})=\boldsymbol{\delta}'\mathbf{x}$ as notation for a linear minimum-mean square error predictor. Then $\mathbf{\Pi x}$ is also interpreted as a linear minimum-mean square error predictor: $\mathbf{\Pi}=Cov(\mathbf{y},\mathbf{x}')V^{-1}(\mathbf{x})$. If we assume that $\mathbf{z}'=(\mathbf{x}',\mathbf{y}')$ is independent and identically distributed across the individuals (exchangeability), then the vector $\mathbf{w}$ formed from $\mathbf{z},\mathbf{zz}'$ is also independent and identically distributed. $\mathbf{\Pi}$ is a non-linear function of $E(\mathbf{w})$, and the sample moments (up to fourth order in $\mathbf{z}$) provide a consistent estimator of $V(\mathbf{w})$. Hence the theory of non-linear regression can be applied to make inferences about $\mathbf{\Pi}$ and the identifiable part of $\mathbf{B}$; the asymptotics refer only to the number of individuals increasing. The relevant theory has been developed by Edmond Malinvaud and

others; we only need to use the special case in which the conditioning variables are constants. Under this approach, it is not necessary to assume that $V(\mathbf{y}|\mathbf{x})$ is independent of $\mathbf{x}$.

Michael Salemi and George Tauchen also consider the bias in regressions of scores on previous scores. They are concerned with the regression to the mean that arises if there is measurement error in the scores. They propose the following clever solution: if a group of M questions are all the same difficulty, so that they all have the same probability (p) of being answered correctly, then the number answered correctly by a given individual has mean $E(y|p)=Mp$ and variance $V(y|p)=Mp(1-p)$. Individuals have different p's, so that $V(y)=EV(y|p)+VE(y|p)=ME(p)[1-E(p)]+M(M-1)V(p)$. Since $E(y)=ME(p)$, we can solve for $V(p)$ in terms of $E(y)$ and $V(y)$; hence the sample mean and variance can be used to obtain a consistent estimator of $V(p)$. When using y as a proxy for p in regression analysis, we can obtain consistent estimates by simply replacing $V(y)$ by $V(p)$ in the moment matrix. (This will not work for general non-linear functions of p; then we need to know the p distribution, which is not identified without restricting it to a specific parametric family.) The analysis does hinge, however, on being able to form groups of questions that all have the same p for every individual. The assumption is testable; if we disaggregate the test scores, the fraction of correct answers in the sample should not differ significantly across the items.

Regulation and Technical Change: Some Largely Unexplored Influences

By GEORGE C. EADS*

Governmentally imposed restrictions on private enterprise are not a new phenomenon. For about 100 years, certain industries, such as the railroads and the electric utilities, have operated under detailed regulation, and even the so-called "unregulated" industries have been subject to antitrust, securities, tax, and labor laws. It is now widely recognized, however, that in the mid-1960's, government regulation entered a new era with the passage of a series of laws aimed at, among other things, protecting the environment, ensuring worker health and safety, and assuring the safety and performance of consumer products.

A number of observers, both within and outside government, have expressed concern that this new use of regulation is fundamentally altering the behavior and performance of *U.S.* private enterprise, with potential repercussions far beyond the intended scope of regulatory activity. (These are reviewed in my referenced paper.) Singled out for special attention has been the fear that the pace and direction of technological advance is likely to be altered.

Historically, technological innovation has been a prime force in economic development. New processes and products have been credited with such diverse benefits as increased employment, increased labor productivity, new opportunities for preventing and curing disease, greater consumer comfort, and improvements in the balance of trade. At the same time, technological development is at least partly responsible for precisely the environmental, health, and safety hazards to which the bulk of the "new regulation" is addressed. The issue is not, therefore, whether regulation affecting the rate and direction of technical change is justified, but what tradeoffs our society is willing to make between the social and economic benefits from further high rates of technological advance and the losses associated with actual and potential new product and process hazards.

A number of recent studies (see, for example, the National Academy of Sciences) have concluded that the regulatory pendulum has swung too far and that certain changes in regulatory processes and procedures must be undertaken if innovation is not to be unduly stifled. I have little quarrel with the general thrust of most of these recommendations—they are sensible and, if implemented, might actually have a positive impact on innovation without unduly hampering the realization of important regulatory goals. However, I would caution against concluding that the topic of regulation and its impact on innovation has been exhausted. Indeed, I am surprised at how little, rather than how much, we actually know. The policy recommendations I have just referred to have been based not so much on an exhaustive body of research, but on a combination of anecdotal evidence plus a measure of common sense. This paper will make some observations about a number of possible influences that regulation may have on the level and character of

*Member, President's Council of Economic Advisers. The views presented here are my own and do not necessarily reflect those of the Council of Economic Advisers. The research on which this paper was based was conducted while I was employed by the Rand Corporation and was funded in part by Rand as part of its program of public service. An earlier version of the paper was presented at the American Chemical Society meetings in September 1978.

innovative activity that the firm may choose to undertake. Though indeed elementary, these observations have not been reflected adequately in the current debate.

Though additional influences undoubtedly could be suggested, I will concentrate on four:

1. Regulation may divert resources that otherwise might be used to fund research.
2. Regulation may change the firm's ability to calculate the payoffs to investments in research and development.
3. Regulation may alter the proportion of benefits that are properly classifiable (from the viewpoint of the firm) as "externalities," and this may change the nature of research the firm is likely to undertake.
4. Regulation may change the optimal institutional patterns for performing certain types of research.

I. Diversion of Resources

This is the aspect of regulation's impact on research activity that has been most widely discussed. Companies that report large shares of their current resources being devoted to responding to government requests for information and to complying with regulations that have already been promulgated infer (if they don't directly state) that some portion of these funds would otherwise have been directed to increased expenditures on "productive" research and development.

The impact of this "resource diversion" argument on the firm's research and development activities would be greatest if, as some claim, the level of resources devoted to *R&D* were generally a simple function of net earnings. Certainly some firms take this "consumption" view of *R&D*. But, if I am to believe the results of those who have studied the innovation process, the vast majority of firms view *R&D* as an investment. In such a case, the resource diversion aspect of regulation affects *R&D* spending not because of its effect on current net earnings, but because it also promises to affect *future* earnings from any new product or process developed as a result of a current expenditure of *R&D* funds. But note: This is equally true for *any* investment the firm might make—in an advertising campaign, in acquiring a new subsidiary, or in expanding production capacity. To the extent that the payoff to *R&D* occurs further in the future than the payoff to other activities, or to the extent that a differential risk premium is assigned to *R&D* projects, *R&D* spending might be particularly hard hit. But the net result on the firm's allocation of future resources across its full spectrum of future opportunities is not all that clear.

Furthermore, in a world of widespread regulation where research can serve both to generate regulatory targets and to provide perhaps the most effective regulatory defense, the short-term payoff to investment in certain types of research may be immense. Take the case of fluorocarbons and the products that make use of them. The *EPA* is currently considering whether the nonaerosol uses of these substances ought to be controlled and, if so, by what means. If all the affected industries are included, tens of billions of dollars in annual sales may be at stake. While the level of costs that any regulatory decision would impose on the economy will properly be a factor in determining whatever action is ultimately taken, the driving factor will be the state of scientific information at the time a decision is made concerning the relationship between the level of emissions of these substances and the state of the ozone layer. Research to develop new uses for fluorocarbons may indeed be in limbo pending resolution of the fluorocarbon issue. But research to investigate the underlying atmospheric relationships, techniques for fluorocarbon conservation, and substitute refrigerants, blowing agents, aerosol propellants, etc., has been greatly stimulated by the threat of regulation. Much of this research may appear (and even may be) "unproductive" from the viewpoint of those firms who might have had other plans for the resources and personnel now involved in it. And it certainly is not the result of any well-planned or logically conceived strategy that identified fluorocarbon substitutes as an area where concentrated scientific inquiry might yield especially promising results at this time. But

inevitably this research will yield new scientific insights and product ideas that can be exploited at some future date.

The one area where *R&D* spending might most be expected to be adversely impacted by the resource-diversion aspects of regulation would be basic research, since in this instance, allocation of funds are often decided on a "rule of thumb" basis. However, as I will observe below, there are reasons to believe that regulation may be acting to *increase* the incentive for firms to engage in what otherwise would be considered "basic research" by changing its character so that it does promise to produce an identifiable payoff to the firms.

II. Changes in the Ability of Firms to Calculate Payoffs from Research

As I have already mentioned, most firms appear to consider their *R&D* expenditures to constitute investments designed to generate future streams of income. As such, their research decisions must meet the test of any investment—the discounted present value of their anticipated future revenues must exceed the discounted present value of their future costs.

However, investments in research and development are characterized by special risks. First there is the risk that the technical objective sought will not be achieved. Then there is the risk that the product or process cannot be profitably marketed. Edwin Mansfield et al.'s work (p. 209) makes clear that the latter are generally more important than the former (due in part perhaps to the small technical advances typically sought in industrial research projects), but that taken together, these risks are high.

Regulation may alter these risks. In determining whether research aimed at producing a new product can be deemed to have achieved "technical success" and the resulting product considered potentially "commercializable," attention must now be given to whether the product can meet both current and anticipated tests for toxicity, carcinogenicity, mutagenicity, teratogenicity, etc. Like the costs of regulatory paperwork, this burden is not too difficult, in principle, for the firm to deal with. The cost of the tests are well known. And, if they are at all "appropriate," the products that are screened out are those that had no business being advanced to the next stage of development. If the tests are "inappropriate," but still serve a screening function that regulatory authorities and the courts consider adequate, only the public is the loser. The firm's profits are likely to be relatively unaffected.

However, a product can meet the conventional standards for "commercialization" and "pass" the battery of tests just referred to and still run into trouble. A corporate decision maker, aware of recent history and knowledgeable of the present somewhat confused state of regulation, must consider the real possibility that, if he proceeds to authorize steps leading to commercialization, somewhere down the road a new and currently undreamed of hazard will be identified that will lead to severe restrictions being placed on the marketability of his product. This may represent the greatest hurdle of all to new product development. For as expensive as the early stages of the product development process (even toxicity testing) undoubtedly are, the real costs of bringing a new product to the marketplace only appear once the decision is made to move it out of the lab. (See Mansfield et al., p. 118.)

It is difficult to know how this latter category of risks can be dealt with by a decision maker. Certainly increased testing of new products is one reasonable response. We are not surprised to learn that at least one company is considering *all* new chemicals to be potentially toxic and is subjecting them to intensive screening (see *Newsweek*, p. 27). However, this response, while laudable, may be insufficient. Accepted standards of toxicity change. Furthermore, certain hazards are so remotely associated with a given chemical that no conceivable test made at the time of the commercialization decision could hope to discover them. The example of fluorocarbons' presently hypothesized impact on the ozone layer is as good as any to cite here.

What, then, is the conscientious executive to do? The most reasonable course would be to employ a heavy additional risk premium when evaluating any product that meets all

presently known tests and that shows strong commercial potential. If the product passes even this hurdle, it may still eventually encounter regulatory difficulties, but the chances are high that it will have repaid its development cost and produced a profit for the company by the time these difficulties emerge. The impact of such a "regulatory risk premium" would be to slow—but not necessarily stop—new product development.

Alternatively, the decision maker might impose a moratorium on commercializing new products. This might prove superior when the regulatory situation is in a state of extreme flux and is expected to stabilize. In such a case, the appropriate risk premium might be so high that the knowledgeable executive, hoping that the situation is temporary, would be hesitant to impose it and thereby weed out new product ideas that might prove acceptable once matters stabilized.

Casual conversations with business executives suggest that behavior of the latter sort may be occurring at present. But absent systematic investigation, I hesitate to give much weight to such statements. Certainly this is a time of extreme regulatory uncertainty. Passage of legislation such as the Toxic Substances Control Act reflect Congress' feelings that *some* form of regulatory screening procedures is absolutely necessary if sensible product introduction decisions are ever to be made. But it remains to be seen whether the legislation will produce any increase in predictability or, more importantly, whether general procedures will emerge that permit existing substances to be classified as to potential threat.

The preceding two categories of impacts have previously been identified and their significance widely debated. The next two I will mention have been much less discussed, but may be of even greater long-term importance.

III. Changes in the Nature of Research that Firms Undertake

Mention of the importance of developing general testing procedures for chemicals brings us to the third of the effects that regulation is likely having on research—its impact on the nature of research that firms undertake.

Most firms normally do not actually engage in much basic research, that is, research whose results are intended to show no apparent applicability to the firm's business. Regulation may however increase the payoff to basic research. Although other examples surely can be cited, the one that comes immediately to mind is the explosive growth of industrial interest in the once obscure field of toxicology.

It is a mistake to assume, as some appear to, that chemical firms gave no thought whatever to the possible consequence to the environment, to consumers, or to workers resulting from the production or use of their products prior to the rise of the "new regulation." Legal doctrines of liability, although perhaps not as strict then as today, provided such an incentive as did concern for the company's public image. But there can be no doubt that the rise of the new regulation and, in particular, certain of its characteristics mentioned earlier, have greatly stimulated toxicological research.

The reasons for the stimulus to basic research reflect the complexity of the incentives presently generated by chemicals regulation. In part they are defensive. Chemical firms, perhaps once content to draw much of their basic research from the universities, now want to stay abreast of research that can overnight direct the regulatory spotlight at important segments of their business. Given the nature of the adversarial process, they cannot wait until a research result is published in a reputable scientific journal and subjected to critical scrutiny by disinterested colleagues.

In part, however, the reasons for interest in toxicology may reflect the power that such knowledge can give a firm vis-à-vis its competitors. Although individual firms will deny that *they* engage in such behavior, it is not unknown for firms to boost their own prospects by suggesting that a competitor's product may embody certain dangers.

Finally, research on toxicology may ultimately generate a level of understanding about the effects of chemicals on biological processes sufficient to permit the develop-

ment of the generalizable testing procedures mentioned earlier. From the viewpoint of the chemicals industry, this would be a great boon, justifying the expenditure of considerable sums of money on research whose immediate applicability appears uncertain.

IV. Changes in the Optimal Institutions for Performing Research

The same factors which alter the *type* of research that the firm has an incentive to undertake also alter the *institutional form* in which that research is carried out. A few years ago, the major chemical firms announced the establishment of an industry-funded institute to undertake research on toxicology (see William Reddig). This represents a sharp break for the chemical industry which traditionally has relied most heavily on in-house research capability. It seems explainable in part because of the shift in what is properly classifiable from the viewpoint of the firm as basic research. But more may be involved. There are strong competitive reasons why it is useful that all chemical firms have equal access to any breakthroughs that might be made in identifying new classes of toxic substances or new mechanisms by which toxicity might be inferred. Such breakthroughs may "clear" entire groups of chemicals or place others in jeopardy. It makes good sense from the viewpoint of the chemicals industry to fund such work cooperatively. I am less certain that it is in the *public* interest, but leave that question to others to debate.

V. Conclusion

Regulation is indeed likely to change both the level and direction of innovative activity in this country. But our knowledge both of regulation and of the innovation process is too primitive for us to tell precisely what these changes will be. Rather than make major changes in the regulatory process on the grounds that they will aid innovation, we are better advised to confine our attention to improving the general climate for innovation and altering those aspects of regulation that even casual investigation may suggest be deleterious. That, by and large, is the direction suggested by most recent important studies; their degree of caution is appropriate.

REFERENCES

G. C. Eads, "The Regulated Society: Implications for Government, Business and the Consumer," remarks before Members of the Town Hall West Forum, Los Angeles, California, June 27, 1979.

Edwin Mansfield et al., *Research and Innovation in the Modern Corporation*, New York 1971.

W. Reddig, "Industry's Preemptive Strike Against Cancer," *Fortune*, Feb. 13, 1978, pp. 116–19.

National Academy of Sciences, *The Impact of Regulation on Industrial Innovation*, Washington 1979.

Newsweek, Aug. 21, 1978.

What Do *R&D* Numbers Tell Us about Technological Change?

By NESTOR E. TERLECKYJ*

Past research involving *R&D* data has produced results which have suggested a number of specific relationships between research and development activities and economic growth, and have helped to identify certain characteristics of the process of technological change. However, many of these results remain ambiguous and their quantitative interpretation and synthesis have been hampered by the absence of definitive data. Many problems surround the concept and measurement of technological change. Until recently, limitations of the *R&D* data precluded analyses of *R&D* inputs in the theoretically preferred form of capital stocks. This paper reviews some of the past research and theoretical discussions and suggests an approach for constructing systematic data which would permit a better focus on technological change than the data now available.

I

The concept of technological change is of rather recent origin in economics, and as yet has not been integrated into the mainstream of economic theory.

Classical economists were conversant with the economic growth effects resulting from improvements in the methods of production in industry and agriculture and with the interactions between growing markets and science applications. But, neoclassical economics and its economic equilibrium formulations tended to exclude the notion of technological change. These theories and the early econometric research in production functions (see Charles Cobb and P. H. Douglas) focused on capital formation, factor substitution, and returns to scale. It was against this background that Joseph Schumpeter brought out his theory of technological innovation and of the economy as a process. He argued that the prevailing neoclassical concept of increasing returns could not account for the observed increases in productive efficiency of the economy. He developed a theory of innovation which simultaneously accounts for growth in productive efficiency and for economic fluctuations. This theory was not integrated into the neoclassical analysis and neither theory led to specification of criteria for measuring technological change. The measures used to represent technical change came from another source.

As Nathan Rosenberg pointed out, it was the series of studies of economic growth conducted at the National Bureau of Economic Research (*NBER*) in the 1950's that generated the contemporary interest among economists in technological change. These studies (Solomon Fabricant, Moses Abramovitz, John Kendrick, 1956, and a later macro-econometric study by Robert Solow) established that the historic rate of economic growth in the United States, and elsewhere, was much greater than could be accounted for by growth in labor and capital inputs. The unexplained difference, or the "residual growth," not accounted for by labor and capital was frequently identified with technological change. Extensive estimates of total factor productivity were prepared by Kendrick (1961). Some economists considered the residual growth to be an error of measurement due to misspecifications of inputs and outputs, or due to omission of inputs not identified in the conventional measures of labor and capital. An alternative approach has been developed for measuring the total factor productivity

*National Planning Association. I would like to thank David M. Levy for helpful comments and suggestions. Research reported in this paper has been supported by Grant No. PRA-76-80754 from the National Science Foundation.

by Frank Gollop and Dale Jorgenson. Data were compiled and published reflecting extensive inclusion of input quantity and quality variables.

Because the interest in technological change originated in the studies of economic growth, research based on index numbers of input and output was given a macro-economic orientation. These studies and subsequent work in growth accounting reduced the unexplained residual either by identifying explainable components (Edward Denison) or by changing the methods of its estimation (Gollop and Jorgenson) but failed to eliminate it. In each case, a residual growth component of about 1 percent a year remained.

II

It is important to realize that the step from the theoretical concept of technological change to the measured productivity residual which sometimes has been taken to represent technological change involves transition from the micro to the macro level, and entails the index number problem in a theoretical and a statistical form. A basic difference exists between the two (except in the one-output/one-input case). Productivity residual is defined at the level of aggregates involving weighted index numbers, while technological change is inherently a multidimensional micro phenomenon.

At the micro-economic level, case studies of innovations have been developed and analyzed by economists for industry (Edwin Mansfield, 1968) and agriculture (Zvi Griliches, 1958). Considerable knowledge has been gained from these studies about the processes of introduction and diffusion of innovations. But a synthesis of the results of the case studies and the results obtained from studies based on productivity growth measured at a more aggregate level has not yet been achieved. Gaps in the available theory and the rather short time which thus far has been available to achieve it may help explain the lack of such a synthesis. The ingredients for a theoretical analysis of new products and product quality change were not available until the 1960's (see Kelvin Lancaster), and these developments have not been linked with a theory of innovation. The empirical work with product characteristics and measurement of quality change in production indices (Griliches, 1961) preceded it by a few years.

Much of the econometric research in this field involved attempts to relate *R&D* expenditures to residual growth in order to identify a component of growth associated with an activity specifically aimed at producing technological changes. Among these macro-econometric studies are industry studies relating research and development expenditures and productivity growth (see Griliches, 1973, Mansfield, and my 1960, 1974 studies) and company studies (Jora Minasian, Griliches, 1980, and M. Ishaq Nadiri and George Bitros). These all established links between *R&D* activities and the residual growth. However, largely because of differences in and limitations of data, a reasonably firm assessment of the magnitude of the technical change component of growth was not obtained as a result of this work. Therefore, the question posed in the title here can be answered only in a limited way and indirectly. Nevertheless, certain specific answers can be suggested and the assumptions under which these answers can be given can be specified.

III

Among the things that can be stated that the *R&D* data tell us are the following five propositions:

1). Technical change can be induced. The pervasive relationship between *R&D* and growth in productivity does not contradict the hypothesis that research and development leads to innovations which result in more efficient production. The pervasive appearance of significant correlations between *R&D* expenditures and productivity growth in studies using different data for different periods and different units (and established within different conceptual frameworks) strongly supports the conclusion that a positive relationship exists and

that productivity increases can be induced by means of *R&D*-based technological innovations.

2). *Ex post* returns to *R&D* investment, both social and private, appear to be high, evidently appreciably higher than returns from fixed capital investment. Also, studies which examined possible effects of *R&D* on real growth of industries (see the author) which purchase inputs from *R&D*-intensive producers suggest that such indirect or external effects of *R&D* investments may be large. Mansfield's case studies (1977), which were based on total cost of innovation, confirm the existence of large external effects at the level of firms.

3). Research and development act as capital. When expenditures are treated as investments and an *R&D* capital stock is derived, more stable estimates of research and development effects are obtained than when *R&D* expenditures are used in such estimations and assumed to be instantaneously depreciated.

4). Government-financed industrial *R&D* has a very different effect on *private* productivity growth, and hence presumably on technological change in industry, than does privately financed *R&D*. Much clearer and more direct relationships have been found between private *R&D* expenditure and productivity growth than for government-financed *R&D*. Econometric research yields indications of positive returns to privately financed *R&D*, and either weak or no indications of comparable returns to publicly financed *R&D*. This, however, may be a problem of data. The time lags involved in the relationship betwen government contract *R&D* and its effects on private-sector productivity, especially for the spillover effects through private-sector applications of government *R&D* in aerospace and electronics, may be much longer than the time lags involved in commercial *R&D* undertaken by firms with a definite business objective. The available time-series of *R&D* expenditures may have been too short to reveal the effects involving the long time lags.

5). There is considerable uncertainty regarding the estimated magnitudes of economic relationships involving *R&D*. Specifically, the rates of return to research and development and the rate of depreciation of *R&D* capital have not been estimated directly. In part, this is due to specification of *R&D*, but it is also due to the problems of measuring economic productivity. Furthermore, *R&D* expenditure is an incomplete measure of the total investment in innovation which is the desired variable. The case studies by Mansfield and John Rapoport establish *R&D* expenditures as an important component of the total investment in innovation, accounting for about 50 percent of the total. But, we are still quite far from having systematically developed industry data on the total investment in innovation and will have to continue to rely on the *R&D* data as a proxy.

There are also a number of problems involving the use of the existing productivity data to analyze technological change components of growth. Some of these problems are discussed in the next section.

IV

In order to arrive at a clearer understanding of the relationship between research and development and, more generally, between investments in innovation and the economic effects of technological change, a number of conceptual and statistical bridges need to be established. These include: a transition from the micro phenomena of technological change, including both inputs and outputs of scientific and technological activities, to their reflection in the aggregate indexes of input and output; between the components of economic growth attributable to technological change and the total changes in the indexes of output, input and productivity; and between data and research findings obtained at the level of the firm, of industry and of the economy which would permit accounting for indirect, external, and spillover effects of innovations.

This, of course, is a huge agenda. Within it, one specific place where a synthesis may be attempted is at the level of product line and establishment-based industry data and analysis. Much of the previous work has

TABLE 1—GROSS AND NET INVESTMENT IN INDUSTRIAL *R&D* CAPITAL IN THE UNITED STATES, ANNUAL AVERAGES FOR MANUFACTURING INDUSTRIES, 1966–73
(Millions of 1972 Dollars)

	Privately Financed *R&D*		Government-Financed *R&D*	
Industry	Gross	Net	Gross	Net
Chemicals	1806	358	51	−48
Textiles	72	13	0	−1
Rubber products	218	47	0	−3
Electrical machinery	1831	315	2101	125
Lumber products	4	−1	0	0
Transportation equipment and ordnance	2732	533	5120	644
Foods	192	39	0	0
Petroleum refining	256	28	5	−1
Furniture	13	3	0	0
Instruments	534	138	295	79
Printing and publishing	1	0	0	0
Machinery, excluding electrical	1684	385	155	−56
Paper	106	16	0	0
Stone, clay and glass products	144	32	4	1
Apparel	8	1	0	0
Fabricated metal products	341	89	272	69
Leather products	0	0	0	0
Primary metal products	237	42	9	−1
Tobacco products	4	2	0	0

Source: National Planning Association estimates based on expenditure data of the National Science Foundation.
Note: Gross investment figures are the product line estimates of privately and government-financed *R&D* deflated by the *GNP* deflator. Net investment data were derived from the changes in the *R&D* capital stocks cumulated from the gross investment data assuming a two-year lag and a six-year straight line depreciation.

been concerned with the relationship between *R&D* expenditures and productivity growth at this level. It is here that the most extensive and consistent data have been assembled and the longest time-series were compiled for both research and development and productivity growth.

I would like to report on some work in progress which is attempting to develop or expand the time-series of industry data on *R&D* and productivity growth. As mentioned earlier, no systematic data on total investment in innovation are likely to become available in the foreseeable future on a regular basis. Hence, we must use *R&D* data and specially prepared data or case study results, and perhaps other technological indicators (such as the patent statistics) to study technologically induced productivity growth. However, we can now improve our use of *R&D* data. On theoretical grounds, net changes in the stock of *R&D* capital were preferred over gross *R&D* investment for use in analytical studies. However, sufficiently long time-series were not available for construction of the capital stock data. By now, the National Science Foundation data for applied research and development by product line has been accumulated for dates since 1958 and its distribution by source of funding has become available in greater detail in recent years. This permits estimation of *R&D* capital stock and of net investment in *R&D*, given assumptions about deflation of the *R&D* expenditure series, and the rate of depreciation of *R&D* capital. A set of gross and net *R&D* investment estimates for manufacturing industries in the period 1966–73 are shown in Table 1.

These tentative estimates of *R&D* capital stock from an ongoing study are now being

TABLE 2—REGRESSION COEFFICIENTS OBTAINED WITH DIFFERENT DATA FOR THE ANNUAL RATES OF CHANGE IN TOTAL FACTOR PRODUCTIVITY, TWENTY MANUFACTURING INDUSTRIES, 1948-66
(t-ratios in parentheses)

		Cost of $R\&D$/Value-Added							
		$R\&D$ Conducted in Industry (1958)		$R\&D$ Embodied in Purchases from Other Industries (1958)		Other Variables			
Data	Constant	Privately Financed	Government Financed	Privately Financed	Government Financed	A	B	C	$\bar{R}^2$
Kendrick 1973	−4.01 (.73)	.27 (2.92)	−.05 (.53)	.81 (3.66)	.12 (.26)	.08 (1.52)	−.04 (3.69)	−.05 (1.37)	.69
Gollop and Jorgenson, 1975	−32.26 (1.79)	.20 (.64)	−.18 (.63)	1.83 (2.53)	1.67 (1.10)	.33 (1.87)	−.03 (.88)	−.02 (.14)	.30
Gollop and Jorgenson, 1977	−5.94 (1.09)	.03 (.31)	.00 (.02)	.07 (.33)	.68 (1.48)	.08 (1.46)	−.03 (2.09)	−.01 (.33)	.28
Kendrick, 1978	−19.26 (1.70)	.31 (1.62)	.08 (.44)	.07 (.15)	.34 (.36)	.22 (2.00)	−.02 (.71)	−.05 (.64)	−.01

Note: Other Variables: A = Percent of Sales Not to Government, 1958; B = Union Members a Percent of Workers in Producing Establishments, 1953; C = Annual Rate of Cyclical Change in Output, 1948-66.

evaluated by means of sensitivity analysis for different depreciation rates, by distributed lag analyses and by reference to other works on depreciation of $R\&D$ capital. Consideration is also given to the use of different depreciation rates for different industries based on decomposition of $R\&D$ between applied research and development and on survey data. However, it may be significant that the data on net investment yield more stable estimates of return to $R\&D$ than the previously used data on gross investment. Thus, in a regression equation of productivity growth estimated for manufacturing industries based on Gollop-Jorgenson data, the $R\&D$ net investment ratio produced a coefficient of 0.7 which was statistically significant at the 5 percent level for the privately financed $R\&D$ and no significant coefficient for government-financed $R\&D$; in the same equation, the gross investment figures did not yield any significant coefficients for either type of $R\&D$ investment.

One of the most difficult issues in the measurement of productivity growth concerns the definition of fixed capital stock. Two different sets of data have been compiled to measure total factor productivity, one by Kendrick (1973) and the other by Gollop and Jorgenson. The Gollop-Jorgenson data contain extensive adjustment of capital stocks for quality which may include technological change components in the capital stock data. Data by Kendrick are also influenced by the assumptions about construction of capital stock. Both of these sets of estimates have been recently revised —although neither series extends to the most recent years. At the industry level, the Gollop-Jorgenson data have been estimated through 1973 and the Kendrick data through 1976. The sensitivity of results to the treatment of fixed capital stock is illustrated in Table 2 which shows the sequence of estimates of the same equation, including the effects of $R\&D$ on productivity growth obtained with different sets of data before and after their revisions. Differences between the data are for the most part due to differences in the fixed capital stock. These differences suggest that, while for other purposes one may wish to proceed differently, for the purpose of analysis of

technological change and specifically of the *R&D* effects on the residual growth, it is essential to be able to isolate the effects of changes in the capital stock which are not related to *R&D*.

To this end, I am now constructing estimates of the residual growth by industry following the approach used by Solow. He estimated the productivity residual as a disembodied shift in a Cobb-Douglas production function by a method involving weighting the capital stock by the share of property income in total income originating. This method can be applied to industry-level data using the latest capital stock data available from the Bureau of Labor Statistics, and possibly even more current estimates. Measures of productivity so derived will be used to estimate growth components attributable to *R&D* along with alternative estimates involving use of capital stock as an independent variable in statistical analysis of changes of output per man-hour. Also, a reconciliation of the results with the results obtained using Kendrick and Gollop-Jorgenson indexes will be performed.

REFERENCES

Moses Abramovitz, *Resources and Output Trends in the United States since 1870*, New York 1956.

Edward Denison, *Accounting for United States Economic Growth, 1929–69*, Washington 1974.

Solomon Fabricant, *Economic Progress and Economic Change*, Nat. Bur. Econ. Res. *34th Annual Report*, New York 1954.

Frank Gollop and Dale W. Jorgenson, "U.S. Productivity Growth by Industry, 1947–73," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement and Analysis*, Nat. Bur. Econ. Res. *Stud. Income and Wealth*, Vol. 44, New York 1980.

______ **and** ______, *U.S. Economic Growth: 1948–73*, forthcoming.

Z. Griliches, "Research Costs and Social Returns: Hybrid Corn and Related Innovations," *J. Polit. Econ.*, Oct. 1958, *66*, 419–31.

______, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, Washington 1961.

______, "Research Expenditures and Growth Accounting," in B. R. Williams, ed., *Science and Technology in Economic Growth*, New York; Toronto 1973.

______, "Returns to Research and Development in the Private Sector," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement and Analysis*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, New York 1980.

John W. Kendrick, *Productivity Trends: Capital and Labor*, New York 1956.

______, *Productivity Trends in the United States*, Princeton 1961.

______, *Postwar Productivity Trends in the United States, 1948–1969*, New York 1973.

Kelvin J. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

Edwin Mansfield, *Industrial Research and Technological Innovation*, New York 1968, ch. 4.

______ **and John Rapoport,** "The Costs of Industrial Product Innovations," *Manage. Sci.*, Aug. 1975, *21*, 1380–86.

______ **et al.,** "Social and Private Rates of Return from Industrial Innovations," *Quart. J. Econ.*, May 1977, *91*, 221–40.

J. R. Minasian, "Research and Development, Production Functions, and Rates of Return," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 80–85.

M. I. Nadiri and G. Bitros, "Research and Development Expenditures and Labor Productivity at the Firm Level," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement and Analysis*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, New York 1980.

N. Rosenberg, "Science, Invention and Economic Growth," *Econ. J.*, March 1974, *84*, 90–108.

Joseph A. Schumpeter, *Business Cycles: A Theoretical, Historical and Statistical Analysis*

of the Capitalist Process, New York; London 1939.

R. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

Nestor E. Terleckyj, "Sources of Productivity Advance: A Pilot Study of Manufacturing Industries, 1899–1953," unpublished doctoral dissertation, Columbia Univ. 1960.

______, *Effects of R&D on the Productivity Growth of Industries: An Exploratory Study*, Washington 1974.

Production Sets, Technological Knowledge, and *R&D*: Fragile and Overworked Constructs for Analysis of Productivity Growth?

By RICHARD R. NELSON*

This paper is concerned with three broad theoretical constructs almost universally applied in analyses of productivity growth: 1) Production sets and their efficiency frontiers (production functions) which sharply delineate the set of input-output possibilities known and available to firms at any time; 2) Technological knowledge, which at any time determines production functions and the advance of which shifts production functions; and 3) Research and development, specialized activities that advance technological knowledge. Economists tend to consider these broad constructs as solidly grounded and universally applicable. In contrast I propose that they are fragile and overworked, and actually misleading for thinking about productivity growth in some economic sectors. More than that, the presumption that they are applicable across all economic sectors obscures exactly those differences which may be the most important ones in explaining cross-sectoral variations in productivity growth rates.

I. Assumptions of Conviction, Interpretation, and Convenience in Orthodox Theory

The assumptions employed in orthodox analysis of productivity growth clearly involve a mix of assumptions, convictions, interpretations, and convenience. I shall argue that the lines between these are very shady. While some of the assumptions are clearly very deeply rooted in convictions, economists tend to think that particular interpretations are the only plausible ones. Hence strong and perhaps justified belief in a basic abstraction tends to carry over to rather dogmatic defense of one particular stylized interpretation of that basic idea. Similarly (strangely enough), certain assumptions that many economists recognize, at least after some deliberation and discussion, as being rather particular interpretations of the more basically rooted concepts or assumptions of convenience and therefore in principle modifiable, turn out to be essential in providing support for and analytic power to the more basic assumptions of conviction; they are integral parts of the analytical package. If they are abandoned, a shadow is cast on the plausibility of some of the more basic assumptions, at least as they are traditionally formalized (interpreted).

The production set concept is clearly a deeply rooted one. It is a basic concept in micro-economic theory. Economists view firms as, at once, having certain capabilities, and as engaged in goal-oriented choice among their capabilities. These ideas are abstracted into the concepts of production set and maximization.

The production set purports to delineate at any time the input-output possibilities that a firm can achieve and which it knows about. Both aspects of the definition are essential to the way the concept is used. The production set of a firm is its choice set of things it can do over which its maximization (usually presumed to be of profit or present worth) proceeds. There are two important presumed properties of the production set. First it is sharp, not fuzzy; that is, there is no vagueness about the input-output opportunities the firm can achieve and that it knows about (although it is admissable within the theory that output be stochastically related to firm inputs, say, as a consequence of uncertain weather conditions). Second, the production set is assumed to be relatively sizeable, certainly including points other than the one the firm presently is

*Yale University.

choosing, and presumably including points the firm never has chosen. These two presumptions might be regarded as special interpretations of the basic idea that a firm has certain capabilities, and does not have others. As I shall discuss shortly, a large sharply defined production set is not the only possible interpretation of that idea. But in analysis of productivity growth considerable revealed conviction resides in the presumption that the production set is sharply defined and not bounded by experience.

Basic orthodox theory is mute regarding why the production set is what it is at any time; the set is a primitive concept. But in the productivity growth literature (and certain other areas of applied microeconomics as well) a certain particular interpretation (theory) of what lies behind production sets has come into almost universal parlance. The presumption is that behind the large (in general infinite) number of points in a production set is a smaller number (likely finite) of feasible techniques. A technique is a known way of doing things that if employed yields predictable inputs and outputs. A production set is generated by imagining all possible levels and variations of use of all possible techniques. Activity analysis is a special model using this technique concept but clearly the basic idea is more general than that, and is consistent with techniques that do not have constant returns to scale and which may admit certain prescribed variation in input coefficients. The question of why the production set is what it is then is answered in terms of the set of known techniques, or (shifting language somewhat) the state of technological knowledge. The shift in language signals another particular interpretation. Technological knowledge carries an engineering connotation. And common usage indicates certain common presumptions about that technological knowledge. The idea seems wide spread that techniques are characterized (at least metaphorically) by blueprints, or recipes as in a cookbook; technological knowledge is in the form of codified how-to-do-it knowledge which provides sufficient guidance so that if one had access to the book one would be able to do it. The "blue print book" idea provides a rationale for the presumption that production sets have sharply defined boundaries. The production set is defined by the set of known techniques, and a technique is either known and in the book of techniques or it isn't.

There is no logical reason why the book of blueprints should be available and known to all, contained as it were, in a public library. But if it were, and access were easy, then one would have a technological knowledge rationale for the proposition that production sets of all firms are the same. The metaphor here is perfectly consistent, of course, with the possibility that different firms have different production sets. In this case, the blueprint book can be regarded as private, with each firm having its own. Presumably there would be the possibility of one firm selling another firm a copy of certain blueprints that it possesses. In principle any firm would be capable of exactly doing anything that any other firm was doing, if it could only have access to the blueprint. An industry production set then could be defined as the points generated by all techniques known by any firm. That all firm production sets are the same then can be regarded as a special assumption of (major) analytical and empirical convenience, but not involving deeply rooted intellectual commitments.

This characterization of technological knowledge is also consistent with a neat, clean characterization of technological advance. Technological advance involves the introduction of new blueprints to the blueprint book, or the replacement of an old blueprint by an improved version. More or better blueprints imply a bigger or better production set, a shift in the production functions.

There is no necessary logical connection between this characterization of technological advance and any particular theory about how technological advance occurs. However, virtually all formal models (there are some exceptions) and almost all of the empirical work aiming to explain differences across industries in rates of measured pro-

ductivity growth have presumed that technological advance is a specialized business. New blueprints get drawn up, tested, and made available to users through a special set of resource-using activities directed specifically to that end—research and development. Technological advance is a matter of conscious choice and investment. Like the "public access to blueprint books" assumption, this one does not seem deeply rooted, but rather more an hypothesis and an assumption of convenience in modeling and empirical work.

These assumptions work together coherently to provide a neat model of productivity growth. The clean-cut production set notion is supported by the book of blueprints notion about technological knowledge. The assumption of a production set common to all firms is a matter of intellectual and empirical convenience, but it has a nice theoretical interpretation in terms of public access to technological knowledge. The assumption of a separate *R&D* activity which expands the book of blueprints supports the idea of a sharp split between moving along a (common) production function and shifting to a new one.

The ideas work together nicely to suggest and justify various techniques of empirical research. Thus if one adopts the assumption that all available techniques have constant returns to scale, and that isoquants are concave and differentiable, and if one assumes as well that firms maximize profits, one can do growth accounting. If one is interested in explaining cross-industry differences in rates of productivity growth, part of the explanation then would reside in differences in rates of growth of various inputs per worker (which in turn might be explained in terms of changes in factor prices and differences in elasticities of substitution) and part by differences in the rate of technological advance as measured by the shift of the production function (which in turn might be explained by the differences in *R&D* spending). Most empirical analyses of differential productivity growth take this form or that of a close relative.

II. Unraveling the Fabric

I have argued above that the assumptions about technological knowledge and about *R&D*, while less deeply rooted than the assumptions about production sets, neatly support not only each other but the basic premises about production sets. In this section I shall relax a few of the assumptions that I think most economists regard as plausible interpretation or convenient for certain purposes, but not essential in the whole logical structure, appealing to certain cases where I think economists generally would agree that they might not be applicable. What may be surprising to some readers is how, as these are relaxed, some of the more deeply rooted assumptions come into question. The purpose of this exercise is not to argue that the canonical assumptions always are inapplicable, but that they are not universally applicable. By assuming their universality one is blinded to some differences across industries that may be an important part of the answer as to why the productivity growth differences are so great.

Let me begin by postulating (or recognizing) that in some sectors not much can be learned from separate *R&D*, and learning is intimately connected with and its effects localized to the environment of "doing." This seems a plausible postulate consistent with at least casual empirical observation. Abandoning the *R&D* theory of technological advance in these cases would not appear to involve abandoning any deep assumptions of the traditional model. But, if one drops the assumption that learning and doing are separate activities, then the clean distinction between moving along a production and shifting the production function is smudged; how the production function shifts depends on where along it one is operating. Also the notion that advance of technological knowledge is something that can be fostered by directing more resources to that end at the least requires a fresh look and more subtle thinking.

The unraveling goes much farther. If the acquisition of technological knowledge oc-

curs through activities independent of applying that knowledge, then the presence of techniques never actually employed within the set of available techniques which generate the production set has an explanation. Absent that separation, there is no rationale that I can see for the presumption that the production set includes points associated with techniques that a firm never has used. Does the idea that production sets are not limited to points associated with already practiced techniques presume a separable learning through research and development activity? If not, what is its basis?

A similar rather unexpected unraveling seems to occur when the assumptions about technological knowledge are relaxed. The "common access to the blueprint library" is the most obviously vulnerable one, applicable as a first approximation in some industries, but clearly a mispecification for some others, say those where secrecy is important. Drop it, and the analysis of productivity growth is complicated by the need to distinguish among firms in an industry and keep track of their relative weights in industry output and inputs.

I propose that the notion that there is *any* library of well-codified technological knowledge at all is equally vulnerable. What if whatever it is that permits a firm to operate a technique in a particular way and with particular outcomes is only in small part describable in a blueprint, or teachable by example, or purchasable in the form of a machine? Then the fact that firm A can operate a particular technique with a particular outcome does not mean that firm B or firm C can, even if firm A helps out their learning in every way it can. The presence of particular and rather special personal talents, or important organizational features, signals that the codified aspects of technique may only be a part of the story. All the "how-to-do-it" books on passing a football and all the expert coaching cannot turn an average young man into a football star. There are no how-to-do-it books or cello teachers that will enable a young cellist to learn to play like Pablo Casals. Good blocking strategies in the playbooks and superior personal talent are only a small portion of what is required for an offensive line to be effective in protecting the passer; the tacit aspects of team cooperation, energy, and stubbornness are an even larger part of the story as almost any coach will testify. A top-flight symphony orchestra is more than a good repertory and strong individual talents, and what it is is not easily imitated.

But all this suggests that in at least the cases above, and perhaps in many conventional industries, the interpretation of that which lies behind production sets as technological knowledge may be quite misleading and unfortunate, at least if technological knowledge has the everyday connotation relating to engineering or hardware. For some techniques personal talents may be important; for others close team work and élan. This is covered over in the conventional interpretation.

But, to the extent that production sets are determined by something other than codified technological knowledge, what is the rationale for presuming that a firm can contemplate and estimate with reliability what would happen if it tried to do things that it has not practiced before? Relaxation of the codified technological knowledge assumption (which at first look did not seem a basic one) seems to have led us to the same place as questioning the learning by *R&D* assumption (also one that at first look did not seem important)—to the proposition that the production set as traditionally characterized may not extend very far beyond the boundaries of current or recent actual practice.

Indeed upon a bit of reflection, the presumption that production capabilities are sharply defined would appear somewhat shaky, absent the codified technological knowledge assumption. Rather than facing a sharply defined set of well-understood techniques with closely predictable inputs and outputs (and an abyss beyond), it would seem more plausible to characterize a firm as having a number of techniques that it can use with considerable confidence,

others which might require a certain amount of *R&D* and learning-by-doing, and still other techniques about which the firm is even more uncertain and which likely would require even more resources and time before the firm could get them under effective control, etc. While this is a plausible (and formalizable) interpretation of the production choices facing real firms in many industries, it is not an interpretation of capabilities in terms of large sharp production sets.

If it is admitted that for some industries at least learning is closely linked with doing, and much that is important about techniques is not codified, then the standard model would appear a poor characterization of the productivity growth processes involved. A model similar to some that Sidney Winter and I recently have been developing would seem more appropriate. Assume that at any time a firm can operate only one technique with real confidence—the technique it currently is operating. Assume that it can try to operate techniques different from the one it is operating, but there is uncertainty about the result of the trial. The predictability of the inputs and outputs it will achieve might be a function of how close the technique it tries is to the technique it currently is using (where closeness can be measured in a number of plausible ways). The predictability of what will happen also might be enhanced by the fact that another firm is using that technique and the firm trying to imitate has access to some information regarding how the imitated firm does it. Add to the foregoing assumptions some additional ones that specify what technique a firm will try to implement, and relate the relative expansion or contraction of different firms to their profit rates. Then one has a perfectly well-defined evolutionary model of productivity growth.

The evolutionary model sketched above and the orthodox model might be regarded as polar extremes along a two-dimensional continuum. As learning is separated from doing, and techniques become better codified, production sets get sharper boundaries, and movements along production functions become more distinguishable from shifts in those production functions. I now go on to propose that perhaps differences among industries in their experienced rates of productivity growth relate to where they are located in these dimensions.

III. The Degree of Technique Codification and *R&D* Specialization—Two Key Variables Behind Intersectoral Differences in Productivity Growth?

My point above is not that the traditional model is a misleading one in all applications, or that an evolutionary model which involves different assumptions is a better general purpose model (although I happen to believe the latter point). Rather, my argument is that sectors and industries differ significantly to the extent in which the traditional assumptions are applicable, and that these differences across sectors and industries are likely to be important reasons behind the observed productivity growth differences.

I am impressed by the extent to which slow productivity growth may be found in sectors where the techniques involve limited codification, and require both teamwork and considerable elements of not easily imitable personal skill. The football team and orchestra examples have much in common with artistic or craft production, and many elements seem to carry over to the provision of services, like barbering or education. These elements militate against the effectiveness of *R&D*.

The education sector is a striking example of the failure to recognize that techniques may have little "technology" behind them, and that separate *R&D* is not very effective. This has led to confusion among economists trying to analyze slow productivity growth in education and frustration among policymakers trying to enhance educational productivity. Economists working on productivity in education have tended to think in terms of well-defined sharp production functions across schools related to some kind of educational technological knowledge. In education, personal elements of teacher style and skill, and the chemistry of the classroom, seem to count at least as much as those aspects of teaching technique

that can be neatly codified, taught, and learned. Policymakers have tried to enhance educational productivity by organizing and funding educational research and development, reasoning that if *R&D* is a powerful new technique-generating device in medicine, *R&D* ought to be powerful in education. But most studies examining the effectiveness of formal education *R&D* have drawn disappointing conclusions.

I propose that the degree of technique codification, and ability to learn through separate *R&D*, while logically independent variables, tend to go together empirically. There are understandable reasons why. If the salient elements of techniques involve special personal skills, or a personalized pattern of interaction and cooperation among a group of individuals in an important way, then one cannot easily infer how it would work from an experiment conducted elsewhere. Research and development is limited to finding effective pieces of technique that can be replicated easily (because they are highly patterned, like a spelling bee, or because they involve physical equipment, like a text book or a movie projector) and which will likely have a positive impact on productivity regardless of the personal skills and organizational structure that must employ it. The targets here may be quite limited. Thus every organization must learn largely on its own, in a somewhat idiosyncratic and inimitable way. This seems to me a situation where the overall pace of learning is likely to be quite slow.

Conversely, if the salient aspects of technique are codifiable, then if a new technique is created once by someone, it is potentially useable by most of the other firms of the industry. Under those circumstances, one might expect that it would be efficient to separate out certain facilities that specialize in learning. Certain types of skill and knowledge might be especially useful when the endeavor is to learn, which are less valuable if it were merely to do. Certain kinds of tricks of the trade, ways to experiment cheaply and reliably, and techniques of modeling or prediction, might begin to be developed. Then *R&D* would emerge as a separate specialized set of activities.

I do not want to be misread as proposing that limited codification and limited power of specialized *R&D* inevitably go together. Industries where *R&D* is very powerful and technological advance is rapid are also likely to be characterized by technologies that are incompletely codified and where personal talents and organizational aspects are important factors determining productivity. In these cases in a sense technological knowledge is changing too rapidly for codification to keep up. But the implications, as in the craft sectors, include considerable and not easily closed differences among firms, absence of well-defined "industry" production functions, and a blurry edge to individual firms' own production sets.

The analysis of the assumptions underlying the traditional analyses of productivity growth could be developed in a number of additional ways. The perspective sketched here leads to consideration of intrafirm organization as an important variable in its own right in influencing productivity, and to some conjectures about how differences across industries depend upon which techniques are governed by hardware versus human organization. Codifiability is likely to relate to the degree of mechanization (the capital-labor ratio) and to the elasticity of substitution between capital and labor in an industry.

DISCUSSION

ROLF PIEKARZ, National Science Foundation: I find my role as the only discussant of these three papers difficult. They cover a great deal of territory and contain many ideas. Also, aside from some minor differences, I very much agree with each paper. As a group, the papers address the major gaps in economic research relating to the market and policy determinants, and to the effects on productivity growth and industrial development of research and development and technological innovation. My major reservations relate to questions I would have liked to have seen addressed.

In capsule form, these papers provide an excellent basis for understanding some important gaps in economic research without making any substantial suggestions to address these gaps. I will focus my attention on the status of economic research in assessing the relationships among research and development, technological innovation, productivity growth, and industrial development. Economic research will be viewed in terms of its contribution to formulating and assessing public policy options. There are a number of public policy domains where the relationships of interest to this session play an important role in discussion and thinking. The three most prominent involve government actions in regulating private sector economic activity (for example, environmental, health, and safety regulation, antitrust actions), policies to stimulate economic growth (for example, tax credits, assistance to individual industries, credits for foreign trade), and government funding of research and development.

TERLECKYJ. My perception of the message of Nestor Terleckyj's paper is that economists using macro data on research and development and productivity growth have not been able to establish clear-cut relationships about the degree to which research and development from various sources and under different circumstances contribute to productivity growth. Terleckyj's paper finds that economists have made an important contribution to our understanding. Notably, they have demonstrated that research and development contribute importantly to both technological innovation and productivity growth.

The Terleckyj paper clearly and succinctly documents that we lack reliable estimates about nearly all of the major elements relating inputs to outputs relevant to thinking about public policy options with regard to research and development and industrial innovation. Terleckyj indicates that only recently have we begun to try to assess the net contribution of research and development activities to the net stock of scientific and technological knowledge. He then points out we have not segregated out the influence of the various types of government research and development funding on private sector technological change and productivity growth. Also, we have no estimates of the contribution of specific classes of scientific findings and technological developments to selected "aggregate indices of input and output." Finally, we have no direct measurements of the relationships between "components of economic growth attributable to technological change" and total economic growth.

In one sense, I find Terleckyj an optimist. Given the above findings, I must disagree with the conclusion that research and development data tell me, "*Ex post* returns to *R&D* investment, both social and private, appear to be high, evidently appreciably higher than returns from fixed capital investment." Without estimates of the above-mentioned relationships, how believeable are the existing estimates about returns from research and development relative to returns from fixed-capital formation? Techniques used to derive existing estimates have a number of limitations and shortcomings whose influences have not been examined or tested out. For example, existing techniques assume known returns to activities complementary to research and development (for example, capital formation) and assume variable returns to research and development. Corporate officials state that many times research and development may be the

outcome of decisions to invest in plant and equipment; under these circumstances, it may be more appropriate to consider returns to investment as a variable. Also, for the preponderant proportion of the studies, the results reflect the outcomes of single equation relationships without considering the possibilities and problems of joint determinancy. Theoretical literature on the topic of technological change, capital formation, and productivity growth has discussed numerous alternative specifications. We must obtain empirical relationships which take into account these possibilities before making broad generalizations about the magnitude of returns from research and development activities.

EADS. The importance of George Eads' paper to economic analysis of research and development rests on his finding that the expressed concerns relating to economic consequences of regulation focus on "fears" of regulation altering the rate and direction of industrial technological advance. He devotes his paper to presenting the case that the policy process lacks information about how regulation influences industrial research and development and technological change. The discussion of the paper is framed in terms of four classes of influence: the amount of resources devoted to research and investment in new or improved technologies; the payoffs calculated by firms for research expenditures and investment in technological change; strategies selected for industrial research (basic vs. applied; short-term vs. long-term) and technological development (new products or processes vs. improvements to existing products or processes); changes in the institutional structure for research and development (cooperative vs. competitive activities) and for commercialization of new technologies.

Overall, Eads has done a good job of identifying the major elements of the public discussion about the influence of regulation on industrial research and technological advance. There is one item I would like to add. There have been some statements that regulation at times tends to diminish social as well as private rates of return from research and development and technological innovation. Proponents derive this conclusion from their observations that regulation reduces long-term and basic research, and diverts investment in technology to modification of known techniques and products. Shorter-term and applied research, and modifications of processes and products, are presumed to have lower social payoffs. Both of the preceding two statements about empirical relationships remain far from confirmed as general propositions for research and development.

Eads' observations about regulation can be extended to the entire spectrum of government policy instruments suggested for enhancing the contribution of science and technology to stimulating productivity, industrial development, and long-term economic growth. Just as with regulation, public discussion tends to give industrial research and technological advance a central role in stimulating *U.S.* industrial and economic progress. Also, policy debates place a high degree of importance on government policy instruments in influencing industrial research and technological advance. Yet, we know little about the magnitude, and frequently the direction, of the influence of various government policy instruments. Instruments most frequently mentioned include lower tax rates (for example, accelerated depreciation) and credits; subsidies (for example, cheaper credit, loan guarantees); assistance for market development (for example, government procurement, export credits); easing antitrust enforcement (for example, cooperative *R&D*); changes in patent provisions, and government funding of commercial research and development.

As an example, during the past two to three years, favorable tax treatment for industrial research and development and for investment in new plant and equipment has been widely advocated and discussed in connection with improving *U.S.* industrial research and technological advance. As with regulation, we lack evidence about four key effects of specific tax measures on expenditures and returns from industrial research and technological development. Notably, by what magnitude would specific tax

measures alter the amount of resources firms devote to research and development? To what extent would tax changes modify the payoffs calculated by firms for potential research expenditures and for investments in technological change? How would tax changes alter the strategy of firms for scientific research and technological development? Where would tax changes influence the institutional structure for research and development, and for commercialization of new technology?

Before concluding my remarks centering on Eads' paper, I would like to make one more point. In order to assess correctly the influence of a particular policy instrument on industrial research and technological advance, it becomes necessary to have the analytical techniques and data to consider the potential consequences of other factors. A major limitation of many past policy assessments has been their inability to separate out the influence of factors operating simultaneously with the policy instruments under examination. Two examples come to my attention. For many years, there has existed a good deal of debate about the degree to which changes in drug regulation adversely affected innovation in the pharmaceutical industry. To an important degree the disagreements revolved around the inability to separate out the influence of the regulations from the effects of the depletion in the stock of scientific knowledge used to improve the technology of drug development. Today, the industry is witnessing a revival in technological development; and we are still unclear about the importance of the rate of changes of the regulatory process relative to recent improvements in the technology of drug development. My second example comes from my position as a funder of research. In many cases, the slowdown in innovation in industries recently subjected to increased regulation has occurred simultaneously with dramatic changes in availability of petroleum and in acceleration of inflation. Proposals for research submitted to us to examine the influence of increased regulations have been subjected to heavy criticisms by peer reviewers because of the inability of potential investigators to demonstrate the capability of their research approaches to separate out the effects of the regulatory changes from the effects of these other factors.

NELSON. From my perspective, the central theme of Richard Nelson's paper is that the prevalent approach used to estimate the influence of research and development on technological advance, and thereby on productivity growth, provides a very limited analytical basis for assessing relationships between research and development activities, technological advance, and productivity growth. This widely used approach assumes that there exist well-defined generally known production sets for the different lines of production. It also assumes technological advance represents the outcome from a separate set of activities to alter the known production sets (for example, investments called research and development). Nelson's paper presents a convincing case for three findings. First, each firm has a limited knowledge of available known techniques. Second, technological advance can and does involve expanding knowledge about a technique through its application (for example, learning by doing). Third, individuals and institutional environments are important to determining how technological knowledge is utilized.

In a more general sense, these findings provide some important insights into thinking about the issues raised in the papers by Terleckyj and Eads. In a broad sense, Nelson tells us that more than research and development must be taken into account when attempting to assess the effect of research and development and technological advance on productivity growth. Specifically, research and development, technological advance, and their contributions to productivity growth must be considered as separate variables. Activities other than research and development contribute importantly to technological advance and contributions of technological advance to productivity growth vary depending upon circumstances. Also, to estimate the influence of research and development on technological advance and productivity

growth, it is necessary to consider the interaction between research and development and other firm activities which contribute to technological advance. Another message transmitted by Nelson is that one should expect the roles of research and development activities and technological advance and productivity growth to differ among industries at various stages in technological evolution and under different institutional and policy conditions.

Concentrating on the policy perspective, more than research and development must be considered when attempting to assess the importance of various government policy instruments fostering or retarding technological advance for specific industries. A major reason for this is that the application of specific government instruments is likely to alter the relative resources or attention firms devote to different activities involved in technological advance. It is to be expected that policy instruments modify individual incentives and institutional conditions.

How Effective have Fiscal Policies been in Changing the Distribution of Income and Wealth?

By MERVYN A. KING*

> We're all living in a house called the Money System; and as a result most of us are suffering from a disease called poverty. (*The Ragged Trousered Philanthropists*)

> When the State spends too much... there is less wealth with which to improve our standard of living. (Conservative Party Manifesto, 1979)

Despite the expansion of empirical research in public finance, there remains considerable uncertainty about the distributional consequences of fiscal policy. For this session, I have been asked to summarize some international comparisons. I shall divide the issue into two questions. How effective has fiscal policy been in reducing inequality? How big are the potential gains from further redistribution?

In Section I, I examine some of the evidence on the redistributive effects of taxes and benefits in the United States, the United Kingdom, and Sweden. I shall concentrate on the distribution among households, and not among tax units, individuals, or by type of factor income. This ignores the fact that the formation of households is itself endogenous and depends, in part, on fiscal policy, especially subsidies to housing costs. Any statement about the impact of taxes on distribution depends on a counterfactual assumption about the distribution which would be observed in the absence of taxes and benefits. Since there is no overwhelming evidence in favor of any one particular set of assumptions, I shall argue that it is helpful to pose a second question, the answer to which does not depend on assumptions about incidence. Given the distribution which emerges from the existing system of taxes and benefits, what would be the gains from attempting further redistribution? Finally, in Section III some suggestions are presented for future research.

I. Fiscal Policy and Redistribution

To evaluate the evidence we must decide on which measure of economic welfare our social ranking is defined. It is conventional to examine income, wealth, or sometimes consumption. In addition, there is the issue of annual vs. lifetime measures, and *ex ante* vs. *ex post* welfare. One of the arguments for schemes of social insurance is that they offer insurance which the market does not or cannot provide, and so reduce the uncertainties facing wage earners in the market place (see Peter Diamond).

All of these issues are very important. But if we are to compare the experience of different countries it is necessary to adopt a more restrictive stance so I shall examine only the distribution of annual household incomes. Table 1 collects together some of the information on the redistributive effects of taxes and transfers in the United States, the United Kingdom, and Sweden. It should be stressed at the outset that there are major

*Professor of investment, University of Birmingham, England.

problems both of interpretation (the pretax distributions reflect the taxes levied) and of comparability (there are differences in the income measure used). Although these are well known, they are no less serious for that. Only for the United States does the income concept come close to "comprehensive" income; for the United Kingdom and Sweden capital gains are omitted.

Nevertheless, the overall picture depicted by Table 1 accords with the conclusions of more detailed bilateral comparisons (see, for example, the study by T. Stark). The distribution of original factor income in the United Kingdom is less unequal (as measured by the Gini coefficient) than in either the United States or Sweden, but after taxes and transfers there is least inequality in Sweden. In both the United Kingdom and the United States most of the reduction in inequality comes not from the tax system but from state transfers (pensions, unemployment benefits, etc.).

This picture is borne out by studies of effective tax rates in the three countries. In Sweden, Thomas Franzén, Kerstin Lövgren, and Irma Rosenberg found that for most household types the net effect of the Swedish tax systems was highly progressive. It is difficult to compare results of this study with those for the United Kingdom and the United States because results are presented only for six income groups and the definition of income includes some transfers but not others. The ratio of taxes to net income rises from 41.1 percent in the lowest income group to 81.5 percent in the highest. For the United Kingdom and the United States it is possible to compare effective tax rates by decile of the income distribution. These have been computed for the United States by Joseph Pechman and Benjamin Okner and by Richard Musgrave, Karl Case, and Herman Leonard using the MERGE file, and for the United Kingdom by the Central Statistical Office. In both cases, of course, the results depend upon the assumptions made about tax incidence. These are difficult to defend, but at least the *U.S.* studies present estimates for more than one set of assumptions which enables us to examine the sensitivity of the conclusions to the assumptions made. In Table 2 these effective tax rates are shown for two assumptions about incidence in the United States, and for the single set of assumptions employed in the *U.K.* study (in which corporation tax and taxes on capital are not allocated to households). Since transfers are much more important in the United Kingdom than the United States (the ratio of government transfers to employees' compensation was about 18 percent in the United Kingdom compared with 9 percent in the United States in the years in question), the effective tax rates for the United Kingdom are shown both for original income and for the distribution by original income plus cash transfers.

The estimates of the effective tax rates depend upon the assumptions made about tax incidence. The numbers in Table 2 suggest that the tax system in the United States is broadly proportional except at the very top, whereas in the United Kingdom it is broadly proportional except at the bottom where it is progressive.

Studies which have examined the distributional impact of public expenditures in the

TABLE 1—GINI COEFFICIENTS FOR HOUSEHOLD INCOME DISTRIBUTIONS IN THE UNITED STATES, THE UNITED KINGDOM, AND SWEDEN

	Original Income	Original Income + Transfers	Net Income
United States 1970	.459[a]	.415	.403
United Kingdom 1972	.415	.348[b]	.318
Sweden 1972	.480	.373[c]	.302

Sources: Okner, Table 7 (average of coefficients for most progressive and least progressive assumptions); Stark, Tables 106, 122, 127.

Notes: Original income is factor incomes before state intervention; the second income concept adds government transfers to factor incomes; the final concept is disposable income after direct but before indirect taxes, except for the United States where the figure is net of all taxes. The *U.S.* data come from the MERGE file and the *Current Population Survey*; the *U.K.* data from the *Family Expenditure Survey*; and the Swedish data from the February Investigation and the *Survey on Relative Income Differences.*

[a]1966

[b]1974

[c]1966

TABLE 2—EFFECTIVE TAX RATES BY DECILE OF INCOME DISTRIBUTION, UNITED STATES AND UNITED KINGDOM

(Shown in Percent)

	United States 1970		United Kingdom 1977	
	Most Progressive	Least Progressive	Original Income	Income plus Cash Transfers
Lowest Decile	18.8	25.8	250.0	3.6
Second Decile	19.5	24.2	53.8	13.3
Third Decile	20.8	24.2	37.5	22.1
Fourth Decile	23.2	25.9	37.0	30.8
Fifth Decile	24.0	26.4	36.6	33.2
Sixth Decile	24.1	26.2	37.4	35.2
Seventh Decile	24.3	26.2	37.2	35.5
Eighth Decile	24.6	26.4	37.4	36.0
Ninth Decile	25.0	26.1	37.8	36.7
Highest Decile	30.7	27.8	38.1	37.3
Average	26.1	26.7	37.6	33.1

Source: United States, Okner, Table 2 and 3; United Kingdom, *Economic Trends*, Jan. 1979, Table D.
Note: The *U.S* numbers were obtained from the MERGE file using the most and the least progressive assumptions about incidence as defined by Pechman and Okner. The number for the lowest decile refers only to the sixth to tenth percentiles. The *U.K.* numbers are based on data from the *Family Expenditure Survey*.

three countries include Musgrave, Case, and Leonard for the United States, Frazen, Lövgren, and Rosenberg for Sweden, and the annual articles published in *Economic Trends* for the United Kingdom. The difficulties in imputing the benefits of public expenditure are, of course, enormous and it is difficult to place much confidence in the results. It is the costs rather than imputed benefits of public services which are allocated to households, and for many items of public expenditure (for example, education and health) the services received by a household depend critically on its age and composition. To examine the distribution of benefits from public services one needs to take account of both the life cycle pattern of benefits received and the differing needs of households of different types. It is, however, instructive, if not particularly surprising, that the three studies referred to above all find that the expenditure side of the budget is progressive. An interesting attempt to relate the effects of the *U.S.* tax and benefits system to the outcome predicted by a model of majority voting on tax and benefit levels is the recent paper by Louis Gevers and S. Proost.

The evidence on the distribution of wealth is even harder to assess than that for income, partly because the data relate principally to the top of the distribution. I would conjecture that the main impact of fiscal policy has been on the share of the bottom 80 percent. Tax concessions have led to a rise in the proportion of personal wealth held in owner-occupied housing which in turn has produced both an increase in the share of the bottom 80 percent and greater differences between owner-occupiers and tenants. In addition, the growth of state pensions has had a marked effect on the share of the bottom 80 percent; in the United Kingdom it rose from 22.4 to 27.4 percent in 1976 if occupational pension rights were included in wealth, and to 44.7 percent if state pension rights were included.

Unfortunately, the data on the distribution of wealth in different countries are not really comparable. Both the United States and the United Kingdom produce estimates based on the estate multiplier method, but in Sweden less reliable data based on other tax statistics are used. For what it is worth, the evidence suggests that the distribution of wealth in the United Kingdom is more

highly concentrated than that in the United States, and that this may be due to the greater importance of inheritance in the United Kingdom. Although there are obvious loopholes in the taxation of transfers of capital in both countries, these seem to be more serious in the United Kingdom. The revenue from transfer taxes actually fell in nominal terms from £459 million in 1972–73 to £396 million in 1978–79. This represents a decline of 65 percent in real terms. As far as the top decile is concerned, the combination of unanticipated inflation and the failure to index the taxation of capital income has been more significant than tax policy.

The effect of fiscal policy on the distribution of wealth is unlikely to have been a "mean-preserving contraction" nor one that can be accurately measured by examining the share of the top 1 or 5 percent. We badly need more evidence on the wealth holdings, and their composition, of the rest of the population.

II. The Gains from Further Redistribution

How big are the potential gains from further redistribution over and above that which results from the existing tax and benefit system? To be more precise, what proportion of national income would we be prepared to give up in order to obtain complete equality of incomes? The answer to this question is Anthony Atkinson's (1970) measure of inequality which is defined by

$$(1) \quad I=1-\left[\sum_i \left(\frac{y_i}{\mu}\right)^{1-\varepsilon} f(y_i)\right]^{\frac{1}{1-\varepsilon}} \qquad \varepsilon \geqslant 0$$

where $f(\ )$ is the frequency distribution of incomes, μ is mean income, and ε is the inequality-aversion parameter. The value of ε may be interpreted as follows. If we attach the same social value to a marginal dollar in the hands of someone with income y as to x dollars for someone with income λy, then

$$(2) \qquad x=\lambda^{\varepsilon}$$

For example, when $\varepsilon=0.5$ one dollar taken from someone on twice-average earnings has the same social value as 50 cents given to someone on one-half average earnings; for $\varepsilon=1$ and $\varepsilon=2$ the figures are 25 cents and 6.25 cents, respectively.

TABLE 3—GAINS FROM FURTHER REDISTRIBUTION (As Percentage of National Income)

ε	United States	United Kingdom	Sweden
0.5	13.7	7.9	7.7
1.0	26.9	15.4	15.8
2.0	49.9	29.4	32.5

Source: My own calculations based on the decile distributions in Pechman and Okner, Tables 4-6; *Economic Trends*, Jan. 1979, Table 11; Stark, Table 102.

Taking the distribution which emerges after all taxes and benefits (but before imputation of the benefits of public services) as a starting point, I have computed the gains from further redistribution (as a percent of national income) in the United States, United Kingdom, and Sweden for three values of ε and these are shown in Table 3. It appears that the potential gains from redistribution are much higher in the United States than in either the United Kingdom or Sweden, and that the ranking of the United Kingdom and Sweden changes as ε increases. This is because higher values of ε pay more attention to the bottom of the distribution, and the share of the lowest decile is higher in the United Kingdom than Sweden although the same is true for the top decile. These gains must of course be set against the efficiency costs of redistribution.

To examine the redistributive effects of fiscal policy we need to compare the degrees of inequality of the distribution of original incomes before state intervention and of final incomes after taxes and benefits. No statistical measure (such as the Gini coefficient) can do this for us and the distributions must be compared using social judgements such as those illustrated in Table 3. But this immediately highlights the impossibility of carrying out such an exercise without explicit consideration of the life cycle. The problem is that in any given year a sizeable number of households have almost no original income and rely entirely on state transfers (such as pensions, unemploy-

ment benefits). In all three countries the share of the bottom decile in original incomes is practically zero. But inspection of (1) shows that if a group has zero income, then for any value of ε of unity or above the degree of inequality is unity, irrespective of the number of people with zero income or the distribution among the rest of the population. Conclusions about the effect of fiscal policy on inequality will be dominated by the position of those with no original income, and this suggests that if we are to use annual data we should focus less on the overall distribution and more on the distribution within a cohort.

Another aspect which has received little attention (but see Atkinson, 1979) is the effect of policy on the ranking of households in the income distribution. Taxes may not only compress the distribution but also alter the ranking of households within it. The former refers to vertical equity, the latter to horizontal equity. Both effects may be important in the eyes of policymakers. In the United Kingdom in 1977, only 31 percent of households in the distribution of original income were in the corresponding decile of the distribution of final income, and no fewer than 27.5 percent of households moved more than one decile (Report No. 7, Royal Commission on the Distribution of Income and Wealth, 1979, p. 62).

III. Conclusions

I have argued that it is impossible to compute useful measures of inequality for the annual distribution of original household incomes, and that, in the absence of improved life cycle data, this makes it difficult to analyze the distributional impact of the tax and benefit system as a whole. A more fruitful task would be to analyze the changes in inequality which would result from particular policy reforms. In itself, however, this is not enough because the ranking of households may change also. Economists' proposals for tax reform often fall foul of politicians because they do not contain an explicit estimate of the distribution of gains and losses implied by the change. To produce such an estimate requires the analysis of individual household data, such as the MERGE file or the *Family Expenditure Survey*, both to investigate the behavioral responses to tax rate changes (allowing for differences in preferences between households) and to incorporate these estimated incentive effects into a distribution of the gains and losses which it is thought would result from the reform being studied. This task (which is one of the aims of the SSRC Programme on Taxation, Incentives and the Distribution of Income directed by Atkinson, Nicholas, Stern, and myself) is less ambitious than that of evaluating the impact of fiscal policy as a whole, but, in the light of the data which are presently available, it may be of more practical value.

REFERENCES

A. B. Atkinson "On the Measurement of Inequality," *J. Econ. Theory*, Sept. 1970, *2*, 244–63.

______, "Horizontal Equity and the Distribution of the Tax Burden," mimeo., Univ. College London 1979.

P. A. Diamond "A Framework for Social Security Analysis," *J. Publ. Econ.*, Dec. 1977, *8*, 275–98.

T. Franzén, K. Lövgren, and I. Rosenberg "Redistributional Effects of Taxes and Public Expenditures in Sweden," *Swedish J. Econ.*, No. 1, 1975, *77*, 31–45.

L. Gevers, and S. Proost, "Some Effects of Taxation and Collective Goods in Postwar America: A Tentative Appraisal," *J. Publ. Econ.*, Apr. 1978, *9*, 115–37.

R. A. Musgrave, K. E. Case, and H. Leonard, "The Distribution of Fiscal Burdens and Benefits," *Publ. Finance Quart.*, July 1974, *2*, 259–311.

B. A. Okner, "Total U.S. Taxes and Their Effect on the Distribution of Family Income in 1966 and 1970," mimeo., Brookings Institution, 1979.

Joseph A. Pechman, and Benjamin A. Okner, *Who Bears the Tax Burden*?, Washington, 1974.

T. Stark, "The Distribution of Income in Eight Countries," background paper no. 4 to Rept. No. 5, Royal Commission on the Distribution of Income and Wealth, London 1977.

Income Redistribution Through the Fiscal System: The Limits of Knowledge

By RICHARD M. BIRD*

In a recent discussion of the welfare basis of real income comparisons, Amartya Sen observed that "...saying less with conviction is not always inferior to saying more with great doubts..." (p. 16), and noted that in making comparisons an important choice must be made between *scope* and *reliability*. Empirical studies of the extent of income redistribution through the fiscal system have clearly opted for scope rather than reliability, so it should not be surprising that the significance of the results of such studies must be regarded with great doubt. Section I outlines some of the main reasons for considering global studies of fiscal incidence in their present form to be of little use. Section II then argues that even if such studies could be satisfactorily carried out they would be of much less value than many seem to think. Fortunately, the principal policy-relevant reasons for such studies can be satisfied by less ambitious but more feasible exercises which, in effect, enable us to say "less" with more conviction, rather than "more" with great doubts.

I. Some Problems with Incidence Studies

Although there is little new in the basic conceptual and methodological arguments against such conventional global incidence studies as those pioneered by Richard Musgrave and Irwin Gillespie (1965), age has staled neither the variety nor the sense of these arguments. Broad studies of tax incidence appear, for instance, to be theoretically impossible because, in Carl Shoup's words, they assume "...what is either untrue or meaningless, namely, that the existing distribution of income-before-tax would remain unaltered if the tax system did not exist" (p. 11). Attempts to estimate the incidence of the entire budget are similarly fatally flawed. This position has been essentially accepted by such recent critics of incidence studies as Lester Thurow, Luc De Wulf, Jacob Meerman, and Charles McLure, and even by some who have themselves carried out full-fledged global incidence studies (see Morgan Reynolds and Eugene Smolensky).

The use of the "no government" counterfactual has recently been justified by Gillespie (1979) on the grounds that "... most fiscal incidence studies contain a *rudimentary* general equilibrium engine which allows for *some* feedback effects on behavioural responses" (p. 17, emphasis added). The argument is that there is no correlation between changes on the uses and the sources side. The assumptions needed to produce this result are strong, though not necessarily wrong. What Meerman calls "RPTO incidence"—or the net effects on factor incomes of budget-induced changes in relative prices, techniques, and output—is most unlikely to be zero, as is in effect assumed in the "no government" counterfactual. Recourse to the "differential incidence" approach of comparing the present system with some other hypothetical system encounters the same difficulties in a less drastic form and thus offers no formal solution to the basic dilemma posed by the nonneutrality of any feasible tax-expenditure system.

Some of the other problems with global incidence studies cluster around the treatment of time. Alan Prest, for example, has noted the apparent inconsistency of the common assumption that personal taxes are not shifted—which implies that factor supplies are perfectly inelastic—and that indirect taxes are fully shifted forward—which implies that factor supplies are perfectly elastic. That either assumption may be appropriate in certain general equilibrium frameworks (see Gillespie, 1979) is not an

*University of Toronto.

adequate response to the criticism that it is inappropriate to add together results that must be derived from inconsistent assumptions. Moreover, the demonstration that a particular assumption is conceptually feasible does not make its use in distributional studies justifiable unless there is reason to believe that it is also empirically relevant. The burden of proof that the results of studies based on such assumptions should be taken seriously might be argued to rest not on the critics of such studies, but on their proponents.

Empirical studies of global fiscal incidence almost invariably purport to depict the effects on income distribution in some particular year. While understandable in terms of data availability, this annual focus has been subject to much criticism as being inappropriate in welfare terms in that it ignores the life cycle nature of the distribution of income. The popularity of the life cycle hypothesis among academics, however, does not mean that lifetime rather than annual income is more relevant (see the book by Enid Slack and myself). Arguments that support the current income focus also cast doubt on the conventional assumption that the appropriate "income" basis for incidence studies is something like the Haig-Simons concept. That is, arguments supporting the annual basis on the grounds that society is allegedly concerned with the distribution of income among individuals over some short time horizon rather than over their lifetimes, seem inconsistent with the choice of an income base for whose welfare relevance there appears to be little popular support. If the analyst is free to choose his own ethically superior income base, why should he not be free to choose the time period he thinks is ethically most relevant? If the choice of time period is constrained by the need to demonstrate that it has some *positive* relevance in the society under study, should not the choice of income base be similarly constrained? Here as elsewhere it is very difficult to disentangle the positive and normative components of fiscal incidence studies.

To mention another time problem, if in the short run the corporate income tax is, for example, assumed to be fully shifted forward, to be borne by all capitalists in the form of a lower rate of return in the "medium" run, and by workers in the form of lower real wages in the long run, what is "the" incidence of the corporate income tax at any point in time? Even if our knowledge were sufficient to permit us to disentangle the web created by myriad past tax changes, what warrant is there for assuming that the mixture of assumptions suitable at one time for one situation is necessarily relevant for a different situation or a different time? More basically, in a sense the question "what is the incidence of the corporate income tax?" is meaningless in any case. As Shoup has stressed "...the very concept of the incidence of a corporation income tax (or other broad-based tax) as such is invalid..." (p. 19) because other changes in the public finance system must take place simultaneously to hold other economic conditions constant so the only effect that can be described quantitatively is the net effect of all the changes.

Great as the problems of tax incidence studies are, they pale in significance compared to those of expenditure incidence studies. Even the terminology is unclear in this field, despite efforts at clarification such as those by McLure and Meerman. To mention only a few points, how should such intermediate "goods" as protective expenditures be treated? Should they simply be disregarded, as Meerman suggests, on the grounds that their benefits cannot meaningfully be allocated among households? Or is the implicit theory of the state underlying this treatment too naive to be credible? If one takes the latter view, what theory of the state should be adopted—a self-interest coalition or median-voter approach, some insurance model, or one of the radical models—and what are the implications of different assumptions?

The treatment of "general" expenditures in fiscal incidence studies in recent years often reflects the ingenious proposal by Henry Aaron and Martin McGuire to allocate the value of public goods to households in proportion to the reciprocal of their marginal utility of private good expenditure, an attempt subsequently made more "operational" by Shlomo Maital, among others.

Most authors who have employed this technique have correctly noted that the empirical basis for such benefit allocations remains inevitably arbitrary because the empirical basis of the assumed utility functions is itself clouded—a situation not likely to be remedied soon in view of the extremely demanding informational requirements of the Aaron-McGuire approach. Further, Geoffrey Brennan has demonstrated that theAaron-McGuire formulation is itself conceptually flawed and that the most appropriate way to allocate the income-equivalent benefits of public goods is, ironically, on a straightforward per family basis.

This very brief catalog of some of the major problems with global incidence studies may be concluded by reiterating 1) that the results depicted in such studies are, as a rule, sensitive to the assumptions underlying them, and 2) that their significance is obscure anyway. Edgar Browning, for example, has recently demonstrated the sensitivity of computed tax incidence results to the reality that government transfers form an important source of income for lower-income groups. Browning assumes that the real value of transfers is unaffected by tax policy, because either the effects of price-increasing taxes are offset by monetary policy, or transfers are indexed. The incidence of the *U.S.* tax system under this assumption becomes progressive at the lower end compared to the regressivity shown in most traditional studies, with the distributional effect of sales taxes in particular being reversed.

What such sensitivity shows is that the results can be no better than the assumptions. Basically, all global incidence studies can do, even conceptually, is to clothe a set of more or less arbitrary and more or less consistent assumptions in a quantitative garb. The air of quantitative precision that results from this procedure is accentuated by the usual tendency to ignore the fact that the estimated variations in fiscal burdens and benefits are often greater *within* income classes than between them.

Even this condensed review of some of the problems with global incidence studies suggests that more has been said and written about this subject than can really be justified on the basis of our limited and imperfect knowledge about the matter we are purporting to study. Indeed, our reach has so far exceeded our grasp in this area that the results of such studies may, on balance, have been more misleading than enlightening.

II. What Can Be Done

This capsule critique of conventional fiscal incidence studies suggests that a more meaningful question than "what's wrong with such studies?" may be "why do we keep doing something so questionable?" A more constructive way to formulate this question might be "what *can* we do that will be less subject to such fundamental doubts and still be relevant for policy"? Fortunately, it turns out that what we can do appears to be more relevant for policy than the global incidence studies criticized above.

Consider why fiscal incidence studies are carried out in the first place. The principal stated purpose is usually to determine the magnitude and direction of income redistribution attributable to government budgetary action. This information permits judgment as to whether government's distributive impact is equitable. There are three issues involved here: Do policymakers want such studies? Do such studies facilitate normative judgments? Do they convey positive information of value?

The fact that public agencies have financed many such studies presumably reveals a desire for them. It is difficult, however, to find any instance in which what policymakers *do*, as opposed to what they may *say*, has been much affected by the "knowledge" about the impact of the fiscal system on distribution displayed in incidence studies. Judging from policy actions, it appears that policymakers are more interested in a breakdown of distributive impacts by region, age, and other such characteristics than in the effects of fiscal policy on income size classes. Similarly, an examination of budgetary policy makes it difficult to believe that the concept of income as a measure of welfare built into incidence studies accords with the real concerns of policymakers. If neither the equity concept

nor the disaggregation provided correspond very well to the concerns of the political system, it is not surprising that the results of incidence studies generally remain in the desk drawer and have little discernible effect on policy—unless, as in the case of the "old" view of property tax incidence, those results happen to accord with what people want to believe anyway. The greatest value of such studies remains what it has always been, as an impressive counterbalance to the undue fascination with the progressive rate structure of the income tax that would otherwise characterize discussions of tax policy in most countries.

A second issue is whether such studies facilitate normative judgments. Gillespie has recently asserted that for all normative questions a comprehensive analysis of budget incidence is needed because "...some such knowledge of the existing state of affairs is necessary before one can calculate to what extent any proposed public policy instrument would alter the existing distribution of income in the direction of the *desired* distribution of income" (1979, p. 11). This position seems rather confused. Normative standards of income distribution presumably relate to the observed distribution, not to some hypothetical "no-government" distribution. The effect of a policy change on the observed distribution may clearly be both interesting and an appropriate subject for study: but there seems no obvious reason why one has to know anything about the distributive effects of the *existing* fiscal system in order to study the effects of such *changes*.

As argued in Section I, all that the present state of economic science permits us to do with confidence is to estimate the short-run distributional impact of small fiscal changes. Even in this case more attention should be paid to the specifics of the economic setting and the accompanying package of policy changes than has been the rule in the past. This is not nearly so great a limitation as one might think, however, since the only policy changes that are seriously considered are generally marginal and, very often, short run in nature. Economists have long deplored the incremental and short-term focus of budgetary decision making. Perhaps those concerned with analyzing the distributive effects of budgetary policy would be better advised to *welcome* such behavior, at least if they agree that this is all we can intelligently talk about at the present time!

No doubt academic curiosity alone will continue to produce fiscal incidence studies. Some of the problems mentioned here—and the many others not mentioned—may be overcome in the future. Others, however, seem unlikely to yield to anything less than a full-fledged, detailed general equilibrium simulation approach—and even then doubts will remain as to what is really being done. As matters now stand, we do not, and cannot, know what the incidence of any fiscal system is. Indeed, the question as usually phrased seems neither meaningful nor directly relevant to public policy. What is needed to improve both the policy relevance and conceptual acceptability of work on this subject is not continued manipulation of masses of numbers in a search for an unknowable truth about fiscal incidence in the large, but rather more detailed study of the distributional impact of specific small budgetary changes in particular circumstances.

REFERENCES

H. Aaron and M. McGuire, "Public Goods and Income Distribution," *Econometrica*, Nov. 1970, *33*, 907–20.

Richard Bird and Enid Slack, *Residential Property Tax Relief in Ontario*, Toronto 1978.

G. Brennan, "The Distributional Implications of Public Goods," *Econometrica*, Mar. 1976, *44*, 391–99.

E. Browning, "The Burden of Taxation," *J. Polit. Econ.*, Aug. 1978, *86*, 649–71.

L. De Wulf, "Fiscal Incidence Studies in Developing Countries: Survey and Critique," *Int. Monet. Fund Staff Papers*, Mar. 1975, *23*, 61–131.

I. Gillespie, "Effects of Public Expenditures on the Distribution of Income," in Richard Musgrave, ed., *Essays in Fiscal Federalism*, Washington 1965.

______, "Taxes, Expenditures and the Re-

distribution of Income in Canada, 1957–1977," unpublished paper, Carleton Univ. 1979.

C. McLure, Jr., "On the Theory and Methodology of Estimating Benefit and Expenditure Incidence," Rice Univ. Program Develop. Stud., 1974.

J. Meerman, "Do Empirical Studies of Budget Incidence Make Sense?," *Publ. Finance*, No. 3, 1978, *33*, 295–313.

S. Maital, "Public Goods and Income Distribution: Some Further Results," *Econometrica*, May 1973, *41*, 561–68.

R. Musgrave et al., "Distribution of Tax Payments by Income Groups: A Case Study for 1948," *Nat. Tax J.*, Mar. 1951, *4*, 1–54.

A. Prest, "The Budget and Interpersonal Distribution" in *The Budget and the Distribution of National Income*, Int. Inst. Publ. Finance, York, England 1968.

Morgan Reynolds and Eugene Smolensky, *Public Expenditures, Taxes, and the Distribution of Income: the United States, 1950, 1961, 1970*, New York 1977.

A. Sen, "The Welfare Basis of Real Income Comparisons: A Survey," *J. Econ. Lit.*, Mar. 1979, *17*, 1–45.

Carl Shoup, *Public Finance*, Chicago 1969.

L. Thurow, "The Economics of Public Finance," *Nat. Tax J.*, June 1975, *28*, 185–94.

The Indirect Incidence of Government Expenditures

By Lester C. Thurow*

Governments affect the distribution of economic resources through four principle channels. They extract taxes, distribute transfer payments and in-kind consumption goods, make human and physical capital investments, and generate market incomes in the process of buying or producing government goods. This latter channel, the indirect incidence of government expenditures, is the subject of this paper.

To determine the impact of government expenditures on the mix of skills, the pattern of industrial demand, and the distribution of earnings, it is necessary to estimate the vector of total government employment and compare it with the vector of what I will call private-private employment. Those private employees who work producing goods or services bought by government are included in total government employment and excluded from private-private employment. The latter is the private employment generated by private demands.

The characteristics of direct government employees are extracted from the *Current Population Survey* (*CPS*) tapes for 1976. The characteristics of indirect government employees are generated using input-output (*I-O*) techniques to estimate the structure of industrial demands produced by government purchases. These industrial demands are then combined with the characteristics of the workers in each industry available from the *CPS* tapes to determine the characteristics of indirect government employees.

*Massachusetts Institute of Technology. The research for this paper was sponsored by the National Science Foundation. I would like to thank Marlys Elavsky, Evan Barrington, Mindy Taranto, Kazuo Ueda, and Jorge Desormeaux for their able research assistance.

I. The Characteristics of Direct Government Employees

In 1976, governments directly employed 18.4 percent of the work force. Of these 19.7 million workers, 14.2 million were state and local government employees, 3.4 million were federal civilian employees, and 2.1 million were federal military employees. Government employed 20 percent of all women, 17 percent of all men, 25 percent of all blacks, and 15 percent of all Hispanics.

If the occupational skill mix (see Table 1) of government is compared with that of the private economy, there is one striking difference. Governments (especially state and local governments) provide employment for a much larger proportion of professional workers than the private economy. While 10.1 percent of all males and 9.3 percent of all females are professional workers in the private economy, the government percentages are 38.5 and 38.1 percent, respectively. As a result government ends up employing 34.5 percent of all male professionals and 49.9 percent of all female professionals in the economy. When one considers all of the female school teachers, social workers, and nurses on public payrolls, the result is not as surprising as it first seems, but it means that highly educated women are highly dependent upon government expenditures for their job opportunities.

Among males, craftsmen and operatives are underrepresented in government (14.8 vs. 33.8 percent). Among women, operatives and service workers account for 40.2 percent of private employment but only 19.3 percent of government employment.

The relative earnings of full-time full-year government workers are shown in Table 2.

TABLE 1—DISTRIBUTION OF CIVILIAN EMPLOYMENT BY OCCUPATION
(Shown in Percent

	Federal Government		State and Local Government		Government		Private Economy		Government as Percent of Total Employment	
Occupation	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Professional	26.5	15.8	43.7	41.7	38.5	38.1	10.1	9.3	34.5	49.9
Managerial	12.2	6.1	12.0	4.0	12.1	4.3	13.3	5.4	11.1	16.3
Sales	.4	.7	.4	.1	.4	.2	6.9	8.5	5.4	.6
Clerical	28.3	65.1	7.8	32.7	13.9	37.2	5.0	31.8	27.6	22.3
Craftsmen	14.3	.3	12.5	.4	13.0	.4	20.6	1.8	8.0	5.3
Operatives Excluding Trans.	2.3	1.8	1.6	.8	1.8	.9	13.2	13.8	1.8	1.5
Transportation Operatives	1.8	1.1	6.3	1.4	5.0	1.4	6.2	.4	9.9	45.0
Nonfarm Laborers	5.4	.7	12.2	.3	10.2	.4	8.6	1.4	14.0	6.1
Private Household	–	–	–	–	–	–	.1	5.4	–	–
Service Workers	8.1	8.3	3.3	18.5	4.7	17.0	10.5	20.6	5.9	16.9
Farmers	–	–	–	–	–	–	2.7	.3	–	–
Farm Laborers	.7	–	.2	.1	.4	.1	2.8	1.4	1.8	1.0
Total[a]	2.2	1.2	6.6	7.6	8.8	8.8	45.7	33.7		

[a]Shown in millions.

TABLE 2—PUBLIC/PRIVATE EARNINGS FOR FULL-TIME FULL-YEAR CIVILIAN WORKERS

	Males			Females		
	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics
Federal Government	115 (99)	133 (117)	136 (120)	144 (124)	153 (130)	155 (132)
State and Local Government	94 (82)	109 (97)	102 (90)	126 (100)	131 (105)	119 (94)

Note: Relative earnings corrected for occupational mix.

Females and minorities are systematically paid a premium relative to their wages in the private economy. Federal premiums range from 33 to 55 percent, and state and local premiums range from 2 to 31 percent. Some of the observed differential can be traced to the differences in the mix of skills employed. The data in parentheses show what government employees would be paid if a correction is made for the occupational skill mix demanded by government. Wage premiums shrink for females and minorities in the federal government and disappear in state and local government.

Table 3 shows data on relative earnings of government employees by occupation. In state and local government large wage premiums exist for male service workers and female managers. Large federal premiums exist for male sales and service workers and for most female workers.

Government may be discriminating less or paying too much, but in either case, government provides a higher wage for females and minorities than is provided by private employers.

TABLE 3—PUBLIC/PRIVATE EARNINGS BY OCCUPATION FOR FULL-TIME FULL-YEAR CIVILIAN WORKERS

	Males		Females	
Occupation	Federal	State and Local	Federal	State and Local
Professional	88	67	111	91
Managerial	105	87	130	123
Sales	132	83	–	108
Clerical	104	86	120	89
Craftsmen	99	84	109	102
Operatives Exclng Trans.	113	83	164	95
Transportation Operatives	84	76	–	–
Nonfarm Laborers	94	85	83	70
Service Workers	127	124	126	105

II. The Characteristics of Indirect Government Employees

Four steps are necessary to derive indirect government employment. 1) Derive the vector of final government demands for 1976 for each of the *I-O* industries. 2) Using that vector and the economy's *I-O* coefficients, calculate the government's pattern of industrial demand. 3) Aggregate the *I-O* industries so that they correspond to the less detailed industrial categories of the *CPS*. 4) Given the characteristics of the workers in each industry and the percentage of each industry employed by government, aggregate to determine the characteristics of indirect government employees.

Since the *I-O* tables do not exist for 1976, the 1970 coefficients were used. While this is less than ideal, there are two reasons for believing that it does not give too misleading a picture. The period from 1970 to 1976 was one of slow growth and the results have been aggregated to the thirteen industries available in the *CPS*. With a high degree of aggregation and a period of slow productivity growth, *I-O* coefficients are probably quite stable.

Overall, 11.2 percent of all private employees are indirectly employed by government (see Table 4). State and local government is a large indirect consumer of construction workers (24.3 percent of their total indirect employment) and the federal government is a large indirect employer of durable goods workers (41.8 percent of their total indirect employment). Both are large indirect consumers of professional workers. As a percent of total private employment, indirect government employment is most important in construction with 30.9 percent of all private workers actually employed by government.

An examination of the distribution of earnings (see Table 5) generated by government reveals that the government provides middle-class jobs while the private-private economy provides more jobs for both low- and high-wage workers. While 47.0 percent of all private-private workers made less than $6,000 per year, only 36.5 percent of all government workers were below this amount. While 4.4 percent of the private-private workers earned more than $25,000 per year, only 3.2 percent of the government workers were in this category. Mean earnings for all classes of government workers were above that for private-private workers with the gap being smallest for state and local workers.

In addition to paying higher wages, the government economy was more egalitarian (see Table 6). The bottom quintile of government workers earned 2.9 percent of total earnings as compared with 2.3 percent for private-private workers. Conversely,

TABLE 4—INDUSTRIAL DISTRIBUTION OF GOVERNMENT EMPLOYMENT
(Shown in Percent)

	Indirect Government Federal	Indirect Government State and Local	Direct Government	Private-Private	Indirect Government as Percent of Private Employment
Agriculture[a]	–	1.4	–	5.4	2.1
Mining	1.3	.8	–	.9	12.2
Construction	3.5	24.3	3.3	4.9	30.9
Durable Goods Manufacturing	41.8	12.9	.6	13.7	16.5
Nondurable Goods Manufacturing	8.2	6.5	.2	11.6	6.9
Transportation and Utilities	10.9	7.2	5.0	5.7	15.3
Trade	9.3	8.9	.7	27.1	4.0
Finance	1.8	2.2	.8	6.3	4.0
Business Services	5.1	4.7	.5	3.8	13.4
Personal Services	3.9	5.5	.1	6.3	9.0
Entertainment	1.7	.4	.7	1.6	
Professional Services	12.6	25.3	49.5	12.7	17.2
Public Administration	–	–	27.7	–	–
Military	–	–	10.7	–	–
Total[b]	3.0	6.8	19.7	77.6	9.8

[a]The absence of federal indirect agricultural employment is misleading. It is produced by the fact that federal government stockpiles of agricultural commodities were being drawn down at the time the *I-O* tables were being constructed. Thus the federal government is a net supplier—not demander—of agricultural commodities.
[b]Shown in millions.

TABLE 5—DISTRIBUTION OF EARNINGS
(Shown in Percent)

Earnings	Military	Federal Civilian	State and Local	Indirect Government	Government (Direct and Indirect)	Private-Private
0–$3,000	2.2	12.4	25.2	24.3	21.8	30.5
3–$6,000	20.1	8.7	15.1	15.0	14.7	16.5
6–$9,000	36.3	13.2	17.9	15.7	17.9	14.9
9–$12,000	17.7	17.5	16.9	18.6	15.9	11.2
12–$15,000	11.5	20.6	10.7	11.4	12.1	8.8
15–$20,000	8.1	15.7	9.5	11.0	10.6	9.5
20–$25,000	2.2	5.7	3.0	4.7	3.8	4.1
25–$50,000	1.9	6.2	1.7	3.6	2.9	3.6
$50,000+	–	.1	–	.7	.3	.8
Mean	$10,670	$12,929	$8,714	$9,367	$9,553	$8,431
Number[a]	2.1	3.4	14.2	9.8	29.4	77.6

[a]Shown in millions.

TABLE 6—QUINTILE DISTRIBUTION OF EARNINGS
(Shown in Percent)

Quintile	Government	Private-Private
Poorest	2.9	2.3
2d	9.8	7.4
3rd	20.3	15.7
4th	25.3	25.4
Richest	41.7	49.1

the top quintile of private-private workers earned 49.1 percent of total earnings while the top quintile of government workers earned 41.7 percent of total earnings.

III. Implications

By comparing the vector of government employment with the vector of private-private employment, it is possible to see how the distribution of earnings would be affected if government expenditures were to decline and private-private expenditures were to expand to hold aggregate employment constant. While a wide variety of tax cuts might be carried out, the simplest assumption is that taxes are cut back so that the vector of private-private demands expands proportionately. If this were done, all those employees who are overrepresented in the vector of government demands would face declining employment opportunities. In these categories relative wages would be subject to downward supply and demand pressures. Conversely, underrepresented workers would experience rising employment and relative wage opportunities.

This gedanken experiment is of course not realistic. If government were to disappear it would simply have to be reinvented under a different name. Private school teachers would replace public school teachers; private police would replace public police. But it does give an idea of what pressures would emerge in the labor market if the country really were to want less of the *types* of goods and services now produced or purchased by government and more of the goods presently purchased by the private economy.

Minorities and women who work in the federal government would face an immediate reduction in their earnings. Regardless of whether their premium is due to less discrimination or to excessive wages, it would disappear when they were forced to shift from government employment to private-private employment.

There would be a major cut in the demand for professional workers, especially for female professionals. Government is a large purchaser of professional services in both its direct and indirect employment. For both males and females there would be an expansion of operative jobs. Looking at the two vectors of occupational employment, there would be major cuts in the demand for educated labor.

Industrial employment would fall in construction, durable goods manufacturing, and professional services, while rising in trade, finance, and nondurable goods manufacturing. Since females and blacks are overrepresented in government they would experience a decline in the demand for their services, but Hispanics are underrepresented and would experience an increase in their demand.

Government is more labor intensive than the private-private economy. In 1976 government purchased 21.2 percent of the domestically purchased *GNP*, but it employed 27.5 percent of the work force. Although the labor-output ratio is higher in government, this does not automatically lead to the conclusion that the capital-labor ratio is lower and that a cutback in government would lower the relative earnings of labor vis-à-vis capital. The capital-labor ratios of the two sectors depend upon the capital stock in each sector. Good estimates of the capital stock do not exist for either sector, but it is possible to look at the flows into the capital stock and gain some idea of the relative magnitudes. The issue is highly dependent, however, on how military investment is treated.

In 1976 the private sector devoted 13.1 percent of its expenditures to plant, equipment, and inventories, while the civilian side of government allocated 12.6 percent of its expenditures to plant and equipment. If these are the correct figures, government workers would bring less than the average

amount of investment with them when they moved from the public to the private sector. If military expenditures are included, however, government plant and equipment investment rises to 17.0 percent. This higher investment rate of the military approximately counterbalances a lower investment rate in other government activities. Thus pressure on the relative returns to capital and labor would depend upon whether government cutbacks were limited to civilian expenditures or whether they included military expenditures.

REFERENCES

Bureau of Economic Analysis, *Input-Output Tables of the U.S. Economy*, Staff Paper No. 27, 1968, 1969, 1970; Staff Paper No. 28, 1971.

U.S. Bureau of the Census, *Current Population Survey*, Mar. 1977 (data tape).

DISCUSSION

ROBERT H. HAVEMAN, University of Wisconsin-Madison: Mervyn King's paper, even considering constraints on length, contributes little to the discussion of the effects of government policy on the income distribution. I find no basis for the conclusion that analysis of the distributional impact of particular policies is more fruitful than evaluating the inequality impacts of the entire budget. The two kinds of analysis serve quite different purposes, but both are valuable in their own right. The case for more limited policy analysis was made by King and is clear; I would in no way dispute it. The case for the more global analysis of the inequality effect of the fisc deserves to be made. It rests, I believe, on three propositions: 1) An overall appraisal of the existing impact of the fisc on equality, even if rough, sets the domain for any policy discussion on the need for, and the desired direction of, proposed tax or transfer reform. It is difficult to argue against better, more consistent numbers when poorer, inconsistent ones would be used. 2) Analyzing the inequality impact of the fisc *over time* using consistent data and a fixed methodology does enable one to separate if only roughly the time-related effect of the fisc relative to demographic and labor market changes, and is not subject to the same critique as are common interpretations of a single year's estimate of fiscal impact. 3) An overall appraisal of the inequality impact of the fisc is in the same league as appraisals of the effect of, say, regulatory policy on economic growth, economic development on fertility, or minimum wages on unemployment —they happen to be the "big questions" about which economists and policymakers alike speculate, and over which they fret. I would not wish to be the first in line to denigrate studies which seek to serve these purposes.

My second comment concerns King's apparent lack of familiarity with recent work on a number of the issues he addresses. He calls for micro-data analysis of the distributional impacts of policy proposals which take into account incentive effects, yet where are references to extensive recent developments in this area? He suggests more work designed to evaluate the social welfare impact of inequality changes attributed to particular policies using explicit social judgements, yet where are the references to the growing body of work involving applications of Atkinson's index in an optimal taxation framework to alternative welfare reform and tax credit proposals? And while he views estimates of the overall effect of the fisc on inequality as relatively unfruitful, there is no acknowledgement of the value of such studies over time and no reference to the most complete and consistent of these time-related analyses.

Finally, I would call into question the general thrust of King's "social welfare effects of incremental reduced inequality" exercise. The *minimum* degree of inequality aversion he admits is that which values a dollar going to a household with twice the mean income at one-half the worth as a dollar to a household with one-half the mean income. Surely there is some inequality aversion in the United States, but I have difficulty believing it to be as large as King's minimum. Clearly his maximum is then quite out of the question. When he uses his minimum value—which may well be a realistic maximum—the permitted inefficiency loss is a meager 14 percent of *GNP*. Evidence on labor supply responses to transfer programs based on both social experiments and cross-section studies suggest that a sizable but by no means enormous (say, $250 billion) tax-financed expansion of the transfer system could well yield national income losses in excess of 14 percent of national income without bringing the nation even close to perfect equality. Indeed, if such a program was a simple expansion in the cost and coverage of the current system of taxes and transfer programs, the reduction in poverty and inequality would be very small and the expected inefficiency losses large. Without major changes in the structure of taxes and transfers, reduction in equality per marginal dollar of efficiency lost is likely to be trivial. The important question,

then, concerns *how* the tax-transfer system is used to reduce poverty and inequality at the margin. Global estimates of the maximum permissable efficiency loss, associated with sufficient nonmarginal expansion of taxes and transfers to achieve complete equality, contribute little to this issue.

PETER MIESZKOWSKI, University of Houston: I agree with a number of points made by Richard Bird, among them the need for analyses which incorporate supply adjustments and a life cycle perspective for some issues, the difficulties involved in allocating or identifying the pattern of benefits of government expenditures by income class, and the need for analyzing within-income bracket variations in the distribution of tax burdens.

But I disagree with a number of points made by the author. One issue is the point first raised by Prest that the assumption of perfectly inelastic factor supplies is inconsistent with the forward shifting of indirect taxes. There is no inconsistency once it is recognized that factor supplies may be quite inelastic to the economy as a whole, but very elastic to specific industries and regions.

A more basic point concerns the use of the "no government" counterfactual. Shoup and Bird apparently believe that global studies are wrong and/or meaningless for they require as their terms of reference an economy in which a tax system does not exist. This appears to follow from the assumption that the existing distribution of before-tax incomes is not changed by the tax system. Although this seems to be the conventional interpretation of existing studies, I believe that it is highly misleading if not incorrect. In the real world a system of lump sum taxation or proportional income taxation is not observed. But by utilizing the concept of differential tax incidence, it seems perfectly natural and correct to interpret the pattern of tax burdens calculated by a global study with reference to a proportional income tax system that yields the same level of real revenue. This approach does not get around the need of calculating the factor price, and factor supply changes with reference to this more neutral tax, but it is a much less exacting and more appealing standard than the hypothetical "no government" income distribution. One practical implication of this standard of comparison is the normative standard which compares the existing tax system, or some component of it, with reference to a proportional tax structure. The policymaker in deciding whether to cut corporate taxes or capital gains taxes makes the decision in part on the degree of progressivity or regressivity of the existing tax system.

Studies of fiscal incidence rarely mix use-of-income and source-of-income considerations. The errors that creep in as a result of this simplification are hard to estimate. The problem in carrying out a fuller analysis are empirical, not conceptual. Sufficient theoretical and computational progress occurred since 1948 so that allocative adjustments in a static general equilibrium sense, and the dynamic intertemporal adjustments in the supply of savings and labor, could be undertaken. But there remains considerable ignorance on key parameter values and it is very difficult, empirically, to link aggregative changes in the overall capital stock to the size distribution of income. There are at present two strands to applied work on fiscal incidence. One strand is simulation of tax effects in a multisector economy with a very limited number of income classes, and more recently simulation work on tax substitution in a growing economy, again with a restricted number of income classes. The other path of analysis is the fiscal incidence studies which deal with detailed income classes, but which are constrained by well-known restrictive assumptions.

The obvious need for, and difficulties in, putting the two pieces of work together may appear to lend support to Bird's arguments that we should attempt to say "less" with more conviction, rather than "more" with greater doubts. But Bird has told us very little as to what we should be doing beyond following an incremental approach in analysis and studying the distributional impact of specific small budgetary changes in particular circumstances. I suspect that if pushed to be more specific Bird would present us with research strategy that would be very similar

to what researchers in the field of fiscal incidence are currently pursuing. The differences would probably center on the issue of small incremental changes versus larger changes relative to the reference point of a proportional income tax. This reduces the issue to whether the errors of analysis are proportional to the magnitude of tax changes. And while errors do unquestionably increase with the magnitude of change, this does not imply that research that moves beyond incremental change is forced into the no tax counterfactual and is fatally flawed as a result.

In the second part of his paper, Bird attempts to argue that even if studies of fiscal incidence could be satisfactorily carried out, they would be of little value to anyone. Although policymakers are interested in the effects of tax policy on specific regions and groups such as the aged, they are nonetheless influenced by studies on the general distributional effects of the fiscal system.

Normative standards of tax policy, especially the politics of tax policy, depend not only on the current distribution of income but also on the perceived level of redistribution through the fiscal system. The corporate tax, despite its many limitations, is justified on the grounds that in its absence the tax system would be insufficiently progressive. The payment of federal grants are justified on the grounds that federal taxes are more progressive than state and local tax systems. A more comprehensive income tax base and wealth taxes are justified on the grounds that the tax system is not progressive at higher levels of income. Fiscal politics and incremental tax changes are influenced, as they should be, by global analyses of fiscal incidence, whatever the errors and ambiguities of these works.

Having attempted to justify the value of global studies of incidence, I am fully aware of the limitation of these studies. A number of the points raised by Bird are on target, but I would like to suggest that we would be better off if the critics spent more of their time gnawing away at ignorance rather than feasting on it.

Economic Development in Historical Perspective

By LLOYD G. REYNOLDS*

The subject of our session is long-term growth in Third-World economies. I interpret these terms in a familiar and conventional sense. The Third World includes all the countries of Central and South America, Asia, and Africa, with the exception of South Africa and Japan. The richer countries I shall call for convenience the "*OECD* group." By economic growth, I mean simply a sustained rise in per capita output. This is not to deny that, in evaluating a particular growth path, it is important to know what outputs were increasing and how the increased income was distributed. But it is confusing, I think, to smuggle these things into the *definition* of growth. By long-term, I mean a period long enough so that growth cannot be dismissed as a temporary boomlet. Simon Kuznets suggests that one needs observations extending at least over three or four decades.

I shall argue that, for many Third-World countries, the development story begins in the second half of the nineteenth century. This is not meant to be heterodox, or even very original. I rely here on the growing number of economic histories of Third-World countries, as well as on the synthetic work of Arthur Lewis, Hla Myint, and Celso Furtado.

When we look back to nineteenth-century experience in a particular country, what are we looking for? There are three historical dating points, each of which might be considered significant for one reason or another: First, we might look for the beginning of a sustained increase in population. Growth in this sense goes back very far. Over the years 1750–1900, W. W. Rostow shows population growing at an average rate of about 1 percent per year in Central and South America, and about 0.5 percent per year in Asia. Data for African countries are fragmentary, but the surmise is that tropical Africa showed little population increase before 1900.

When population is growing there is a presumption that *GNP*, and especially the critical component of food supply, is growing at least at the same rate. There are cases of population growth accompanied by immiserization, but these are the exception rather than the rule.

A second date, and the one on which I shall focus, is that at which one observes a perceptible increase in output per capita. Most of us hesitate to speak of "take off," which implies a smooth and rapid ascent to an unchanged cruising altitude. I prefer Rudolf Bicanic's formulation that nations *creep over the threshold* of economic development. It is usually several decades before one can be sure that the threshold has been passed. This is what I call the period of *growth acceleration*. By the end of this period, a developing economy has become a mature economy in the sense that a positive growth rate—fluctuating, to be sure, over the course of time—has been built into the structure of the economy.

My taxonomy of economies thus contains three categories: *stagnant* (in terms of per capita output), *developing*, and *mature*. Such a classification is more useful than the conventional tables arraying countries in order of per capita income in a particular year. Taking a long view, *rates of change* are more significant than *levels* at one point in time.

A third date, to which some attach special significance, marks the appearance of modern large-scale manufacturing. But experience both in the *OECD* group and in Third-World countries suggests that a country usually develops for quite some

*Yale University.

time before this milestone is passed. The agricultural sector is typically more central than the industrial sector in early economic growth. Industrialization is better regarded as a normal consequence of sustained growth rather than as a source or criterion of growth.

My prime concern, then, is with correct dating of growth acceleration, country by country, throughout the Third World. I shall comment briefly on three time periods: 1870–1914; 1914–45, and 1945–73.

I. The Golden Age: 1870–1914

The modern growth procession, led by the United Kingdom and its earlier followers, had by 1870 been joined by most other European countries as well as the areas of overseas settlement in North America and Australia. Japan joined the procession in the 1880's. The *GNP* growth rates in the *OECD* group in the late nineteenth century, though they may today appear modest, were much higher than anything experienced before that time. Kuznets shows a median growth rate of around 3 percent per year in total output, and around 2 percent per year in output per capita, over the years 1870–1913.

This would in itself have been a strong stimulus to international trade. But a further stimulus came from rapid reduction of transportation costs, particularly ocean freight costs. The steamship had largely replaced sail on passenger routes by 1850, on freight routes up to 3,000 miles by 1870, and on long-distance freight routes by 1890. Freight rates fell dramatically. The rate from Java to Europe, for example, fell from around 70 shillings a ton in 1870 to 20 shillings in 1910. The opening of the Suez Canal in 1869 was a major event, at one stroke bringing Asia 4,000 miles closer to Europe.

Railroad building also helped to knit Asia and Africa into the expanding international economy. Railroads running inland from the ports, while designed partly to move troops rapidly to trouble spots, also served to carry manufactured goods inland and to bring plantation crops and other export products to the ports. By 1913 India had about 34,000 miles of railroads; China, 6,000; Indo-China, 2,500; the Dutch Indies, 2,000. Railroad development in Africa came relatively late, as did colonization itself, but some 9,000 miles of railroads were in place in tropical Africa by 1913.

The result of all this was a rapid expansion of tropical exports, which reached a crescendo in the thirty years preceding 1914. W. Arthur Lewis estimates that from 1883 to 1913 tropical exports grew at almost exactly the same rate as industrial production in the advanced countries—about 3.6 percent per year. Agricultural exports grew a bit more slowly than this—about 3 percent per year—but this was offset by a higher growth rate for minerals and other exports.

The impression of rapid growth which one gains from overall statistics is confirmed as one looks at the economic history of particular countries. I have not yet surveyed all of the country literature, and there are countries for which no good economic history exists. But thus far, in Latin America, I find convincing evidence of growth acceleration in Argentina, Brazil, Chile, Colombia, Mexico, and Peru. In Asia, there is similar evidence for Burma, Ceylon, Indo-China, Malaya, and Thailand (but not for China, India, Indonesia, and the Philippines). The African countries, mainly at this time colonies, were latecomers to economic development. Several of them show rapid export growth from 1900 to 1914, but this short period does not permit one to speak of sustained growth with any assurance.

This export growth was accomplished mainly by mobilizing unused land and unused labor time. Most countries at this time had ample land reserves. In many countries, too, peasant farmers were in a position of "subsistence affluence," in which conventional necessities could be produced by working less than a full day or a full year. Export production could be added on, with no sacrifice of subsistence production, simply by a reduction of leisure time, compensated by increased consumption of imported consumer goods. The classic examples are the West African cocoa farmers and the rice farmers of Southeast Asia. In other cases, immigrants provided the necessary addition to labor supply, as in Argentina, Brazil, Chile, Ceylon, and Malaya.

Why did some Third-World economies climb onto the growth escalator during this period while others did not? Land availability was certainly a factor. In Egypt, India, Java, and China, land scarcity inhibited development. In some other countries, including Venezuela and the Philippines, land was concentrated in the hands of large owners and was not available to smallholders. In addition to land, many tropical crops require either heavy, guaranteed rainfall or irrigation facilities. In general, wet lands fared better than dry lands in the export boom. Governments also differed widely in the speed with which they moved to create roads, railroads, ports, and other necessary infrastructure; and this was just as true of independent governments as of colonial governments.

To what extent did the output growth of this period raise income for the ordinary man? This depended partly on the method of production of export products, whether they came from mines, plantations, or peasant smallholders. Peasant production was most likely to lead to broad-based income growth. At best, growth benefited only those—perhaps a quarter to a half of the population—who were drawn into the money economy. The majority remained locked into subsistence production, and one could sometimes find groups, such as artisans out-competed by foreign imports, who suffered an income loss. But the fact that income gains were unequally distributed should not lead one to deny that gains did occur.

Perhaps a more important question is to what degree the growth of this period helped to lay a foundation for subsequent growth. As Lewis points out, "...what matters in the long run is not how large today's income is, but the extent to which the proceeds are used to make possible self-sustaining growth at a reasonable level. This involves what is done to create a productive agriculture, infrastructure, a modern elite, and financial capacity."

Here the record is quite mixed. Governments of this period, no doubt influenced by laissez-faire economics, typically undertaxed and underspent. They did reasonably well in creating physical infrastructure. But with few exceptions they neglected education, and did little in the area of agricultural research and experimentation. They sometimes trained local administrators and civil servants, but did little to develop business entrepreneurship. They were generally indifferent or hostile to industrialization. Only India, Ceylon, and a few Latin American countries had made a start on industrial development by 1914.

To attribute all this to colonialism is simplistic. The development record of colonial governments ranges from quite good (American, Japanese) to quite bad (Spanish, Portuguese). But the independent countries of Venezuela, Ethiopia, Turkey, Iran, Afghanistan, and Thailand were not exactly models of growth-oriented government.

II. The Interwar Years, 1914–45

World War I brought great loss of life and capital, and broke down national boundaries and customary trade patterns. Before this damage had been repaired, the Great Depression struck a massive blow at the world economy. The effects on world trade were disastrous. Between 1913 and 1938, trade grew at only 1 percent a year, and the percentage of world output entering into trade *fell* substantially. Tropical exports suffered along with exports in general. In addition, the terms of trade turned against primary products relative to manufactures, further reducing capacity to import in the Third-World countries.

The result was that countries which had been growing quite rapidly before 1914 now grew more slowly or stagnated, while at the same time conditions were not propitious for additional countries to embark on growth acceleration. I expect that, as I continue to investigate this period, a few cases of growth acceleration will turn up. Korea and Taiwan are clear cases. Japanese colonial administrators worked systematically and effectively to turn these areas into rice baskets for Japan. The infrastructure development, increases in agricultural productivity, and manufacturing growth achieved before 1940 laid a foundation for the rapid growth of these economies after 1945.

Other possible cases include Turkey and Iran, both of which underwent political revolution in the early 1920's and installed regimes interested in modernization and economic growth. A few of the African economies, too, which had begun to grow around 1910, may by 1930 have been well launched on sustained growth.

III. The World Economic Boom of 1945–73

I end this period with 1973 because I believe it will turn out in retrospect that the long postwar boom ended at that point, and that the world economy entered a new and bleaker period. Further, the events of 1973–74 created a sharp division of Third-World countries into oil producers, which are able for the time being to develop with virtually unlimited supplies of capital, and non-oil producers, whose growth prospects have been reduced by rising energy costs.

From 1945–73 the world economy grew as never before, confounding the prophets of stagnation who had flourished in the 1930's. In the *OECD* countries, *GNP* grew at an average annual rate of about 5 percent, and *GNP* per capita at close to 4 percent. These rates are well above those for any previous period of comparable length. World trade grew even faster than world output, aided by continued reduction of transport costs and also of tariff barriers. While the hub of world trade continued to be an interchange of goods among the *OECD* countries, demand for imports spilled over to Third-World countries as well. Their export volume grew over the period 1948–72 at an average annual rate of 8.1 percent. There was probably a mild downtrend, however, in terms of trade for primary products as against manufactures. This export performance casts doubt on the contention of Nurkse, Prebisch and others that trade is a less-powerful engine of growth today than it was in earlier times.

In the more successful Third-World economies, output grew a bit faster from 1945 to 1973 than it did in the *OECD* countries. The top fifteen Third-World countries show a *GNP* growth rate of about 6 percent per year, even after adjusting for upward biases in the data. But two footnotes should be added. First, even in these successful Third-World economies, *GNP per capita* rose more slowly than in the *OECD* countries because of substantially higher rates of population growth. Second, viewing the Third World as a whole, the dispersion of national growth rates was greater, and the median growth rate considerably lower, than in the *OECD* countries. This accounts for the familiar fact that the income gap between the First World and the Third World has been widening.

Which were the fastest growing Third-World economies in this recent period? Omitting city states and looking only at countries of substantial size, the list includes: in Asia, Turkey, Iran, Pakistan, Malaysia, Thailand, Taiwan, the Philippines, and South Korea; in Latin America, Brazil, Colombia, Venezuela, and Mexico; in Africa, Kenya, Nigeria, and the Ivory Coast. In an analysis of these countries published two years ago, I observed that their growth did not consist in spectacular performance in some sectors offset by poor performance in others. Speaking broadly, they show good performance across the board. First, with few exceptions, they raised food output at better than 5 percent a year, fast enough to feed their growing populations at a higher level of nutrition; and the exceptions were cases, such as Iran and South Korea, where food imports were paid for by the rapid rise of non-food exports. Second, they raised manufacturing output at about 10 percent a year, following the standard progression from light to heavy industry; and by the end of the period manufacturing in most countries provided 15 to 25 percent of *GNP*. Third, the median growth rate of exports for the group was only 4 percent in the 1950's, but this rose to 7 percent in the 1960's and to more than 15 percent from 1965 to 1973. Fourth, governments showed marked ability to raise domestic revenue through taxation. Figures on current revenue as a percentage of *GNP* show a median of about 10 percent in 1950–51; but by 1971–73 this had risen to 20 percent.

Why did these countries do so well while others did less well? Size of country,

resource base, political leadership and administrative competence, the accident of being closely associated with one or more of the richer countries—these and other factors must have entered in. But in addition let me advance a "head start hypothesis:" One reason for rapid growth after 1945 was that growth was already under way well before 1945. Note that the success stories are concentrated in Southeast Asia and Latin America, which were also in the forefront of the 1870–1914 export boom. In almost every country on the list, growth acceleration can be dated from the late nineteenth century or from the interwar period.

IV. Concluding Comments

The hypotheses which seem reasonable to me at this point can be summarized as follows:

1. The correct dating of growth acceleration, country by country, is a good deal earlier than has sometimes been assumed. Many Third-World countries experienced growth acceleration in the favorable environment of 1870–1914, and a few did so even in the unfavorable years 1914–45. It is wrong to think that, after growth accelerated in Japan in the 1880's, nothing significant happened until after 1945. This presumed gap in the growth procession is imaginary.

2. Third-World countries which have had a high growth rate since 1945 are in general countries whose growth began in earlier decades, typically in the late nineteenth century. There certainly are additional countries whose growth acceleration began after 1945, and which by 2000 will correctly be classified as mature economies. But it is still a bit early to be sure which countries these are.

3. In both the nineteenth and twentieth centuries, the growth of Third-World economies has been closely related to growth in the *OECD* countries, and to international flows of goods, finance, and technology. Growth in the world economy provides, as it were, an escalator onto which Third-World countries can climb, and onto which many have climbed at differing points in time. But other countries have not succeeded in doing so. There clearly are conditions within a country which can prevent economic growth. Just what these are remains rather mysterious.

Comparative Indian Economic Growth: 1870 to 1970

By ALAN HESTON AND ROBERT SUMMERS*

Because of the colonial relationship of India and England, the subject of long-term Indian economic growth has generated partisanship and controversy. As Daniel Thorner has asked, where can one find a more dramatic presentation of conclusions than that of William Digby, who had imprinted in gold on the spine of his book the per capita income of "*Prosperous*" *British India* in 1850 as 2d., 1882 at $1\frac{1}{2}$d., and 1900 as less that 3/4 d. And Digby's adversaries were persons of moment like Lord Curzon of Kedleston who had claimed that per capita income in India had risen between 1880 and 1900 (see also F. T. Atkinson).

I. Long-Term Estimates of National Income Growth

While the temper of this controversy has long since subsided, many of the questions raised are important and relevant to contemporary research. One question that concerned many students of India was its level of income relative to England. John Crawfurd in the 1830's, for example, wishing to convey the conditions in India to a British audience, compared the earnings of rural laborers in the two countries. He concluded that while the wages of rural labor in India were Rs. 15 to Rs. 20 per year, necessities cost one-third as much in India as Britain. Therefore, their wages in terms of necessities would be 45 to 60 rupees, or in English terms 4.5 to 6 pounds a year, which was still substantially less than the wages of an English laborer.

In his remarkable contribution to the study of Indian national income, Dhadabhai Naoroji was also interested in comparing the per capita income in India and England, but with a particular concern for demonstrating the higher burden of taxation in India. He placed the per capita income of India at Rs. 30 about 1870 compared to that of Rs. 450 for England. His estimate is of great interest both in addressing the question of the absolute level of incomes in India, and in placing the poverty of India in comparative context.

Recent long-period studies of Indian national income include those of M. Mukerjee, S. Sivasubramonian, and K. Mukerji, and modifications and extensions of these works by Angus Maddison and Heston. Despite disagreements about the treatment of specific sectors like agriculture, government, and other services, the extent of difference in the various estimates are comparatively modest. For example, it is possible to defend estimates of per capita income which would show no gain, or a small decline in per capita income from 1870 to 1970; on the other hand, estimates of 0.5 percent per annum growth in per capita income from 1870 to 1970 seem on the high side in recent analyses of historic Indian growth. These differences, particularly between negative and positive trends in per capita income growth provide rich materials for the rhetoric of those evaluating the impact of British rule in India. However, in terms of Indian economic performance vis-à-vis Europe, Japan, or the United States, the entire range of estimates of Indian performance would suggest a substantial relative decline, *primarily* because of the rapid growth in these other world areas.

II. Some Special Features of Indian Economic Growth

Within this pattern of slow overall growth of per capita *GDP* there are several clear

*University of Pennsylvania. We are indebted to Irving B. Kravis and Raaj Sah for comments on an earlier draft.

developments of note. Not surprisingly, per capita growth was uneven: the 1871–1911 growth rate was about 0.7 percent per annum, 1911–46, virtually no growth, and from 1947 to 1970, about 1 percent per annum, using Heston's series from 1870 to 1947, which is similar to Mukerjee's. Population growth on the other hand, was very small from 1870 to 1920, and then began an acceleration that continued until 1970; the totals for undivided India 254 million in 1870, 305 million in 1920, and for India and Pakistan, 668 million in 1970. It is difficult to judge if the 1911–46 stagnation in per capita income is entirely real, though there are certainly depressing factors, like, for example, declining growth of the British market, and the rapid reported-population growth after 1921. With respect to the role of population, however, Morris Morris has pointed out that the apparent decrease in food availability per capita in India predated or coincided with the increase in population (and improvement in health standards) after 1920, so the role of population growth is hardly clear. But as shown by John Wall, there is no tendency in the same period for the price of food crops to rise relative to other goods, which one might expect from some models of the role of Malthusian population pressure pushing on limited land.

Two other developments during the period involving the integration of the different regions are also of considerable interest in the development framework. There is much evidence suggesting that economic integration as measured by the reduction of price differentials, increased substantially after the introduction of the railroad, and particularly in the three decades from 1870 to 1900. Further there would appear to be very uneven developments among the regions, with the areas of Eastern Uttar Pradesh and Bengal in particular showing very little growth from 1900 to Independence, and from 1947 onwards as well. Any complete explanation of Indian economic history in the last century must deal with the heterogeneous experience of several very large regions with populations in the tens of millions that have shown aggregate growth not so much lower than many industrial countries, while at the same time dealing with several relatively stagnant regions.

III. Comparative Indian Growth

The major point examined in the remainder of this paper is the consistency of various estimates of long-term Indian economic growth within a comparative framework, where comparisons are made with experience in the United States. The contemporary experience of India and the United States is used as a benchmark from which to examine Indian historic growth in the light of likely structural changes in the two economies during the past century. While we believe our use of growth rates of national income in current and constant prices in this illustration is defensible, the reader is advised to interpret the empirical evidence as suggestive. If the results from this exercise seem useful, their empirical validity should then be examined more intensively for India and other countries.

Growth of real per capita income in constant prices in India (undivided India prior to 1947, and Republic of India after 1947) appears to be from 1.00 in 1870 to 1.75 in 1970. For the United States for the same period the index may have moved from 1.00 to 5.70 implying a relative decline in India to 31 percent (1.75/5.70) of its 1870 standing in relation to the United States. Does this seem consistent with the levels of income in the two countries in the two periods? Two sets of level comparisons are in Table 1 that seem to bear on this question. The purchasing power comparisons involve many sources of error. (This is especially true for 1870, because the 1870 comparison is based on only twenty price comparisons of goods and ten salary comparisons.) However, there would not appear to be any systematic direction of error in the purchasing power comparisons. Three purchasing power estimates are presented for 1870, each based on a particular assumption about the weight of services and the relation of priced and nonpriced services. No claim can be made that these necessarily bracket the true figure.

TABLE 1—PER CAPITA INCOME IN INDIA AS A PERCENT OF THE UNITED STATES

On the Basis of:	1870	1970
Exchange Rates	9.1	2.1
Purchasing Power Parities	25.1*	6.9
*Range of Estimates	20.7 to	29.4

The first point we would make is that the relative national growth rates for the two countries are consistent with the real product comparisons in the second row. That is, if we take as a benchmark the 1970 estimate that Indian *GDP* was 6.9 percent of the United States (Irving Kravis, Heston, Summers, 1978a, p. 10), then extrapolating this backward using the above national growth rates would put India at 22.5 percent of the United States in 1870, which is not far from our best guess estimate of 25.1 percent.

In both years the purchasing power estimates put India far higher than comparisons based on exchange rates, for familiar reasons that are most explicitly treated below. If one were to apply the same extrapolation to the exchange rate estimate for 1970, it would imply that India was 6.8 percent of the United States in 1870, in contrast to our contemporary estimate of 9.1 percent at exchange rates of 1870. In what follows, we would like to illustrate for a very simple world the pattern of structural change that might give rise to the India–*U.S.* experience just outlined and then to examine its relevance for understanding national growth rates and past and present income differences.

IV. An Illustration of Comparative Growth

We begin with a received view of what leads to divergences between estimates of income differences from exchange rates and purchasing power parities. A more systematic description is provided elsewhere (Kravis, Heston, and Summers, 1978b, p. 219). In this model, it is assumed that there are traded and nontraded goods, with exchange rates dependent on the former, so that the price of traded goods will, other things the same, tend towards equality. However, nontraded goods include many services and commodities, like construction, and these sectors are relatively more labor intensive. Furthermore, in these sectors growth in labor productivity is usually slow. At a point in time relatively affluent countries have high output per worker in traded goods. If there is substantial mobility of labor between traded- and nontraded-goods industries in each country, opportunity cost of workers in traded goods industries must be met by comparable wages in nontraded-goods industries. Then because labor productivity is lower in these sectors, the relative prices of nontraded goods may be expected to be higher in more affluent countries.

This same cross-section phenomenon tends to occur over time. As countries grow, the productivity in the traded goods industries will tend to increase, with the result that the relative price of traded goods will tend to decline. In the example given in Table 2, we have tried to illustrate this kind of cross-section, time-series, scenario, with numbers that very approximately fit the India–*U.S.* experience. Lines 1–3 give the prices and quantities of traded and nontraded goods for both years for both countries, and the current price output. National output is calculated at 1870 and 1970 prices in lines 4 and 5, and the Fisher Ideal index—their geometric mean—of the 1970-1870 output figures are given in line 6. The relative growth of India and the United States in national terms roughly corresponds to the actual estimates discussed above, though, of course the price and quantity figures are purely illustrative, for no data are in fact available on the traded and nontraded breakdown for 1870. It has been assumed that quantities of nontraded goods rise in both countries between 1870 and 1970, but not as rapidly as the quantities of traded goods.

The relative prices of nontraded goods in India and the United States have been changed (increased) to match the assumed price movements given in the model above. For India the relative increase in the price of nontraded goods has been assumed to be

TABLE 2—ILLUSTRATIVE EXAMPLE OF ALTERNATIVE MEASURES OF INCOME GROWTH AND REAL INCOME DIFFERENCE

	India				United States			
Part A. National Comparisons	1870		1970		1870		1970	
	P	*Q*	*P*	*Q*	*P*	*Q*	*P*	*Q*
1. Traded Goods	8.80	20	40	37	8	85	10	600
2. Nontraded Goods	2.45	20	17	33	6	85	17	400
3. Output at Current Prices	225		2041		1190		12800	
	(1)	(2)	(3)=(2/1)		(4)	(5)	(6)=(5/4)	
4. Output at 1870 Prices	225	406	1.81		1190	7200	6.05	
5. Output at 1970 prices	1140	2041	1.79		2295	12800	5.58	
6. Fisher Index			1.80				5.81	

Part B. International Comparisons of Output	1870 India	1870 U.S.	1870 India/U.S.	1970 India	1970 U.S.	1970 India/U.S.
	(1)	(2)	(3)=(1/2)	(4)	(5)	(6)=(4/5)
7. At Indian Prices	225	956	.235	2041	30800	.066
8. At *U.S.* Prices	280	1190	.235	931	12800	.073
9. Fisher Index			.235			.069
10. At Exchange Rates						
A. Using Price of Traded Goods		Exchange Rate=1.10	.172		Exchange Rate=4.0	.040
B. Using 1970 Relation of Exchange Rate to Relative Price of Traded Goods		Exchange Rate=2.08	.091		Exchange Rate=7.5	.021
11. Exchange Rate Deviation from *PPP*		1870			1970	
A. Line 9/Line 10A		1.37			1.72	
B. Line 9/Line 10B		2.58			3.28	

much less than for the United States because of the relatively slow growth in output per worker in traded goods, particularly agriculture, and the probable growth in disguised unemployment.

With the information in lines 1 and 2 of Table 2, it is then possible to make international comparisons between India and the United States in both 1870 and 1970, valuing each country's output in the other country's prices. These comparisons are given in lines 7 and 8, with the ratio of output of India to the United States being given in column (3) for 1870 and column (6) for 1970. The geometric mean or Fisher index for each year is given in line 9, and corresponds roughly to the historic estimates in Section III.

In the last line of the table, the exchange-rate comparison is presented, where as discussed above, we assume that the exchange rate is determined by the relative price of traded goods in each year. Since traded goods are relatively cheap in the United States, the more-so in 1970, India relative to the United States would look much poorer at exchange rates than in terms of purchasing power parities. The exchange rate comparison in line 10A does not produce as low an estimate of India-*U.S.* as the basic figure in Section III. The reason for this is that the model assumes that the exchange rate is the ratio of the prices of traded goods in the two countries, whereas as measured in empirical work it is much higher.[1]

[1]The reason for this is that the classification of traded and nontraded goods is necessarily imperfect. For example, in the 1970 measure, items like perishable fruits and vegetables are treated as traded goods, even though they have a high nontraded component of distributive services in their price. This has the effect of lowering the measured price (the rupee-dollar ratio in our example) of traded goods relative to the exchange rate. Similarly nontraded goods may have a traded component as when carpenters from India work in Saudi Arabia, or citizens of Kuwait have medical treatment in a Bombay hospital.

The example has been constructed so that the quantity comparison in line 10A at exchange rates computed at the relative prices of traded goods bears the same relation in both years to the actual exchange rate comparisons as reported in Section III, and reproduced in line 10B. Put another way, the ratio of the exchange rate to the relative price of the traded goods is the same in both years in the example. The remaining comments in the paper will use the comparisons of line 10B with line 6, but the qualitative results also could be inferred from line 10A.

In Table 2 the decline in relative income computed on the basis of the exchange rate is more than implied by national growth rates because the relative price of services is rising faster in the United States than in India, and this is not offset by relative declines in the quantities of nontraded goods. The exchange-rate comparisons illustrate a point that is obvious, perhaps, but worth emphasizing; exchange rates can be very misleading for international comparisons not only at a point in time, but also over time for the same reasons. The problem is that they are dependent on only a fraction of the products and services entering into *GDP*. A comparison between two countries at different points in time based upon the exchange rate will not properly reflect relative growth rates if the relative composition and prices of traded and nontraded goods change. This is illustrated in the example of Table 2 and in the historical estimates of Section III.

A more important lesson to be drawn from the illustration is that national growth rates (constructed by some average weighting procedure) do reflect real output growth that is comparable over time for each country and across space. That is, relative national growth rates will reflect relative changes in real income between countries, where real incomes are based on purchasing power parities. While this point may also appear obvious, it has very significant implications for judging the gaps between countries.

V. Conclusions

To illustrate the point just made, consider the comparison of India and the United States for 1870 which is 9.1 percent at exchange rates, or an eleven-fold gap. If this difference were to be closed in fifty years with the United States growing at 1 percent per capita, it would have required India to grow at 5.8 percent per capita, whereas to close the estimated real gap of 4 to 1, under the same circumstances would have required a 3.8 percent growth rate per capita. (In 1970 the same calculations would imply a 9.2 percent growth rate at exchange rates, and 6.6 percent at purchasing power parities (*PPP*).) Clearly the perception of the extra resources required to equalize incomes between countries are very different, and very much less when real incomes are compared, which would seem the appropriate measure in contemporary or historic studies.

Another feature of the India–*U.S.* comparisons is that apparently the differential between exchange rate and *PPP* estimates of relative income has increased. This is consistent with contemporary cross sections where the difference between *PPP* and the exchange rate is greater the greater the difference in real income between the countries; since the gap in real income between India and the United States grew, so would the gap between the *PPP* and the exchange rate. This change has occurred, judging from contemporary cross sections, because over time relative quantities of traded and nontraded goods have changed less in both India and the United States, than relative prices. If the prices of traded goods declined more rapidly relative to nontraded goods in the United States than in India, it would be consistent, assuming no offsetting quantity changes, with the increasing differential found between the exchange rate and *PPP* of the rupee between 1870 and 1970.

There are several directions for future study suggested by these kinds of comparisons. First, comparisons between countries at two different points in time provide a kind of check on the underlying national

series for both countries. As consistent historic estimates are built up for a number of countries, it would provide a benchmark to which other countries could be linked. Second, as contemporary knowledge about the relationship of exchange rate and purchasing power estimates increases, it should provide us with insight into how to integrate data on wage rates and relative prices in analyses using other measures of aggregate economic activity.

REFERENCES

F. T. Atkinson, "A Statistical Review of the Income and Wealth of British India," *J. Royal Statist. Soc.*, June 1902, *65*, 209–72.

William Digby, *"Prosperous" British India: A Revelation from Official Records*, London 1901, p. 534.

John Crawfurd, *A Sketch of the Commercial Resources and Monetary and Mercantile System of British India, with Suggestions for their Improvements by Means of Banking Establishments*, 1837. (Reference is to K. N. Chaudhuri, ed., *The Economic Development of India Under the East India Company 1814–1858*, Cambridge 1971, pp. 228–29.)

A. Heston, "National Income," in *Cambridge Economic History of India*, Cambridge forthcoming, ch. 4.

Irving B. Kravis, Alan Heston, and Robert Summers, (1978a) *International Comparisons of Real Product and Purchasing Power*, Baltimore 1978.

——, ——, **and** ——, (1978b) "Real GDP Per Capita for More Than One Hundred Countries," *Econ. J.*, June 1978, *88*, 215–42.

Angus Maddison, *India and Pakistan Since the Moghuls*, New York 1971, p. 166.

M. D. Morris, "Towards a Reinterpretation of 19th Century Indian Economic History," *J. Econ. Hist.*, Dec. 1963, *23*, 606–18.

M. Mukerjee, *The National Income of India, Trends and Structure*, Calcutta 1969.

Dhadabhai Naoroji, *Poverty and Un-British Rule in India*, Delhi 1962, p. 53.

Daniel Thorner, "Long Term Trends in Output in India," in his and Alice Thorner, eds., *Land and Labor in India*, Bombay; New York 1962.

J. Wall, "Foodgrain Management: Pricing, Procurement, Distribution, Import and Storage Policy," in *India*, Occas. Paper, World Bank Staff work. paper no. 279, Nov. 1978, p. 84.

Brazilian Development in Long-Term Perspective

By ALBERT FISHLOW*

Brazil is now the eighth largest market economy in the world. Its 1979 per capita income of over $1,500 in current dollars, its sophisticated industrial structure, and its recent dynamism combine to give it a prominent place among the advanced developing countries. Not surprisingly, recent Brazilian experience has been an important subject for the analysis of economic development, not all of it uncritical.

Such a focus on the Brazilian "miracle" and its aftermath is too limited. Brazilian growth and industrialization have a long history as do the regional and personal income inequalities that today increasingly command attention. Understanding the origins of Brazilian growth in the nineteenth-century export economy and the subsequent patterns of import substitution will not yield better rules for short-term debt management or monetary policy. But the longer perspective makes clear that sustained economic development requires more than following the dictates of the international market and getting prices right. Indeed it may on occasion involve rejecting such signals.

I shall interpret four significant episodes of Brazilian expansion since the last century that illuminate a complex and changing stimulus to domestic growth in the periphery of the extension of the world economy:[1]

1. The impact upon Brazilian growth from integration into the expanding international economy of the nineteenth century came late and in diluted form.

2. Adverse shocks of war and depression in the twentieth century reinforced import substitution and contributed to a rise in the share of manufactures but at the expense of efficient expansion of capacity.

3. Post-World War II economic growth can be explained better as a consequence of accelerated inputs of capital and technology than by emphasis upon import substitution.

4. The recent "miracle" has built upon, rather than supplanted, earlier industrialization, and has depended as much upon integration into international capital as trade markets.

The net result of Brazilian adaptation to the opportunities and limitations posed by the external market has been a continuous growth of per capita gross product since the beginning of the twentieth century at rates clearly exceeding those of the nineteenth. There is evidence of acceleration in industrial growth beginning with the Great Depression; more rapid agricultural expansion since 1945 also seems to have occurred. That record, in the face of the wide swings experienced in world trade, is an impressive accomplishment. It testifies to the prevalence of an inherently pragmatic rather than ideological commitment of Brazilian economic and commercial policies over the long term.

I

Brazil's emergence as a major producer of coffee had come by the mid-nineteenth century; indeed, by the 1830's coffee accounted for more than 40 percent of the value of total Brazilian exports. Through the third-quarter of the century, although trade as a whole grew quite satisfactorily—the income terms of trade at 3.8 percent between 1848–52 and 1873–77—coffee export volume showed only modest advance, 1.9 percent. This period's favorable trade results were the consequence of special circumstances: a new rise in coffee prices to unprecedented levels; a last surge in sugar exports; and the favorable but temporary

*Yale University.

[1]Much of the historical discussion can be found in my referenced article. For other, and sometimes differing views, see among others, Annibal Villela and Wilson Suzigan; Nathaniel Leff (1969, 1972, 1973); Carlos Palaez and Wilson Suzigan.

effect of the U.S. Civil War on cotton exports. Increased receipts did not represent a permanent response to the quickened pace of industrial development in the core; nor was there significant change in the traditional, slave plantation economy.

During the decade of the 1880's, while Brazil's export proceeds almost stagnated, there was an annual increase in coffee export volume of 5.7 percent, the result of the new São Paulo coffee economy. Its development took a shape reminiscent of patterns in other areas of recent settlement. Extension of railways, construction of port facilities, and rapid population growth via immigration all characterized this regional expansion over the course of the quarter-century before World War I.

This belated response to new opportunities had only limited effect upon Brazilian development as a whole, however. The decline of sugar and cotton in the northeast—in which more than a third of the population resided—did not release much manpower to the dynamic south. Ironically, São Paulo showed net internal immigration before abolition of slavery, not thereafter. Rather a deliberate program of subsidized recruitment of immigrants—especially from southern Italy—was relied upon to satisfy manpower requirements. Domestic reallocation in the face of changing international comparative advantage failed—opening a gap in regional incomes and rates of growth that was to be aggravated through the next half-century.

A second feature of this late São Paulo response was its limited scope for continued coffee export growth. Within a decade, it became apparent that serious problems of oversupply presented themselves. The price of coffee experienced ever-wider oscillations —falling in the 1890's from a sterling price of £4.09 in 1893 to £1.48 in 1899 as new Brazilian capacity came into production.

The reaction was twofold. One was an early interest in control of supply to regularize prices; the Taubaté agreement in 1906, supported by foreign loans, stands as the first such effort. The other consequence of limited coffee export growth—the annual increase in volume between 1883–87 and 1909–13 was 2.9 percent—was an early and continuing incentive to reallocate export profits to other activities.

What stands out therefore in the Brazilian case is the limited applicability of the "classic export engine of growth" view. Regional immobilities meant that changed comparative advantage brought with it significant and growing internal disparities. Limited export possibilities inspired early efforts in São Paulo to diversify and insulate the economy from external fluctuations. Monetary heterodoxy at various times broke inflexible linkages with the gold standard, with the exchange rate and its variations consequently playing a central role in influencing domestic economic activity.

II

Despite early efforts at import substitution in the 1890's and subsequent rapid growth of manufactures between 1906 and 1912, Brazil on the eve of World War I had only a primitive industrial structure. It continued to import textiles, relying upon external sources for more than a third of consumption. Although considerable advance in substitution for processed foodstuffs had occurred, agricultural imports remained large, rivaling those of iron and steel. The domestic capital goods sector was essentially nonexistent. No more than 3 percent of the labor force was engaged in manufactures.

The war interrupted this development process based upon simultaneous expansion of exports and import substitution, both of whose labor requirements were met by immigration. Trade was curtailed, especially imports, and population flows were much lower. There was irregular and slower industrial growth; explained largely by import substitution, a reduced rate of capital formation imposed by limited access to imports, and evidence of rising profits partially inspired by inflation.

The war was therefore not a starting point for industrialization, but neither was it an unequivocal check. It reversed the rise in imports that seemed to be gathering force immediately prior to the war. (Abundant

foreign exchange generated by the Amazon rubber boom had made external purchases attractive.) It was a source of windfall gains that should have found their way into the considerable postwar expansion of industrial capacity. It came at a time when the industrialization was still fragile. As late as 1919, imports of manufactures probably constituted almost as large a contribution to industrial supply as domestic value-added. The threat of external competition remained real, despite high-tariff protection.

The reality of that threat became apparent in the decade of the 1920's. Conservative domestic monetary policy had as its objective revaluation of the milreis. Industrial growth slowed under tight credit and cheaper imports. Between 1922 and 1926 it stagnated, recovering thereafter. Had the influx of imports not occurred—financed in part by foreign loans—industrial growth in that same period would have been close to 8 percent. The decade as a whole ended with excess capacity and uncertain direction.

The Great Depression definitively resolved both. Brazil was among the countries initially hardest hit, not least because new investments in coffee in the 1920's began to bear fruit only in the 1930's. The price of coffee fell by 60 percent between 1929 and 1931, and continued at low levels for the rest of the decade. Capacity to import remained at two-thirds its late 1920's magnitude.

Yet Brazil succeeded in sustaining a growth rate in industry of more than 10 percent annually between 1932 and 1939; that of gross domestic product attained almost 6 percent. It did so in the first instance by sustaining internal demand despite the external shock. Demand was stimulated by the policy of protecting, at least partially, the internal income of coffee producers, through the purchase and destruction of coffee stocks, as well as through real devaluation; by the federal government deficits, not by any means initially intended; by the increasing receipts of the coffee sector (modestly recovered between 1932 and 1936); and by an expansive monetary policy.

The role of the coffee sector should not be exaggerated in all of this; it was not much more than 10 percent of income, and its receipts could not be fully sustained. In the absence of compensatory policy, however, the results would have been less positive. It was possible for Brazil to proceed as autonomously as it did because the economy was not so open: the ratio of exports to gross domestic product was not much more than 15 percent. That made a compensatory policy feasible.

Supply responded initially on the basis of the previous excess capacity accumulated in the 1920's. Competitive imports fell rapidly as limited foreign exchange was diverted to debt service. Labor intensity of production increased during the decade as extra shifts were employed fully to utilize scarce capital. New investment was undertaken during the period, but in amounts limited by access to imported machines and concentrated in the newer sectors.

The rapid growth in manufactures during the decade was underwritten by a massive substitution of imports. In the aggregate, about half of the observed increase in value-added in industry can be explained by a reduction in import supply. The industries that grew most rapidly were the intermediate and capital goods sectors, a reflection of the earlier industrialization in the consumer goods sectors of textiles, shoes, clothing, and foodstuffs. Industry became increasingly concentrated in São Paulo, which before the depression had an industrial structure more oriented to the new and technologically advanced sectors.

By 1939, Brazil's manufacturing sector employed some 9.5 percent of the labor force. It had diversified appreciably beyond its predepression state. Industry remained, however, relatively backward. Import substitution by necessity, while an impressive adaptation, imposed a capital-scarce industrialization less capable of competing efficiently with other countries. It required, as well, a labor force that could be absorbed at approximately constant real wages—a possibility that could be sustained for short periods under special economic and political

circumstances, but not over the longer term. Industrialization under the impulse of adverse shocks is not without its limitations.

III

The Post-World War II industrialization evolved in a context of deliberate policy that differentiates it from earlier periods. Three elements characterize the disequilibrium model that emerged. One was a reliance upon commercial policy to effect a large sectoral transfer of resources between agriculture and industry. Early on, an overvalued exchange rate taxed agriculture while subsidizing industry. Later, however, direct advantages to industry through more favorable exchange rates had to be fiscally financed as export rates rose.

That necessity imposed a new and disturbing feature of rapid growth in the 1950's: accelerating inflation. Governmental expenditures were not matched by increased revenues despite an enhanced public role in supplying the infrastructure required by continuing modernization. The deficit rose and was financed through accommodating monetary expansion.

The third characteristic was a reliance on foreign capital. Direct investment in the dynamic industrial sector satisfied entrepreneurial and technological requirements of a more sophisticated and diversified development of manufactures. The consumer durable and intermediate sectors favored by the government could not have expanded solely under domestic auspices. Foreign capital also financed the progressively more pressing needs for foreign exchange as export receipts stagnated after 1953.

All three elements in the model were subject to reversal. Exchange overvaluation eventually took its toll on the diversified supply of exports, and not merely bonanza coffee profits. Forced savings relied upon a continuing rise in inflation, and such acceleration provoked its own distortions; as inflation mounted to 40 percent, resistance mounted. Foreign investment, supplemented by private short-term capital inflows attracted by very favorable terms, were not stable and dependable sources of finance. Real capital formation was itself discrete, and not a smooth process; short-term capital was volatile and subject to early repayment.

Efforts to resolve these potential contradictions in time proved inadequate. Exports responded slowly to more favorable policies, and confronted a world market in which prices of primary products were falling. Fiscal and monetary policy were controlled irregularly and inadequately. Foreign capital dried up, provoking continuing balance-of-payments crises in the early 1960's and partial and unsuccessful liberalization.

It is now fashionable to criticize this flawed import substitution industrialization from the perspective of the successful more open development strategies of the late 1960's and early 1970's. Three comments on the Brazilian experience, at least, place that strategy in a more favorable light. One is that the contribution of import substitution per se during the 1950's was not a significant component of demand for manufactures. The ratio of imports to domestic production was already quite low in 1949 and the subsequent decline during the decade accounts for less than one-fifth the observed growth of manufactures. This is in contrast to the Great Depression experience. Even a constant import-income ratio would have involved a significant margin for continuing development of manufactures.

In the second instance, Brazilian industrialization was not as inefficient as the inflated tariff structure has made it appear. Impossibly high rates of efficient protection are the result of temporary measures to deal with the balance of payments, and do not measure real cost differentials. The subsequent economic miracle had its basis much more in the diversified industrial structure and excess capacity derived from the past than divine mystery. How can Brazil's subsequent export performance in manufactures be explained—a scant few years after import substitution industrialization had

been dismissed? The capital-intensive modern industrial development was in fact accompanied by a 2.4 percent annual rate of productivity growth that far exceeded previous experience.

Thirdly, more realistic exchangerate policy to encourage exports encountered real difficulties in international markets. It is wrong and ahistorical to ignore that the record expansion of world trade in the 1950's did not favor primary products. Brazil in particular faced competition from lower-cost tropical competitors. Policies and priorities correctly favored industry under such circumstances, and were not as prejudiced against simultaneously exploiting external market opportunities as some retrospective views suggest.

IV

The Brazilian miracle post-1964 was linked to a new economic strategy. that new model was characterized by greater integration, partially subsidized, into international markets; by a larger and more centralized fiscal capacity; by a structure of subsidies and incentives favoring profits rather than wages; by enactment of monetary correction to diminish inflation-induced distortion; by institutional reform of the social security system, internal financial markets tax laws, etc; and by technocratic economic management as a counterpart to authoritarian political control.

For all its frank commitment to capitalism as the source of capital accumulation, the model never corresponded to a free enterprise prototype. Brazilian economic strategy has been more pragmatic and rooted in an interventionist tradition. Government participation in the economy, an object of criticism in 1963 *and* in 1979, increased after military intervention. Public investment, directly in infrastructure and through state enterprises, rose as a percentage of capital formation. Regulation of economic activity did not wither away. Subsidies and incentives proliferated and price controls as well; they were accepted and indeed welcomed so long as they were consistent with rapidly increasing profits. Public control over resources, via taxes as well as the forced savings accumulated through the social security system, expanded.

Nor did the emphasis upon industrialization significantly alter. Agricultural exports did not receive anything like the same subsidies as manufactures, and indeed continued to be taxed by the system of protection against foreign imports. Agricultural production for domestic consumption received little attention, not surprisingly since the principal foodstuffs are produced more than proportionately on small- and medium-sized units. Foreign investment and modern capital intensive technology were welcomed.

The model has been praised for the extraordinary growth it fostered between 1968 and 1973: a rate of aggregate expansion in excess of 10 percent a year is no mean achievement. It has also been criticized for its failure to distribute income and opportunities more equitably. Both are appropriate, and neither needs further elaboration here.

Rather I wish to stress the special character and importance of Brazilian integration into world capital markets as a condition of success of the model. Despite rapid and unprecedented export volume growth—about 10 percent since the mid-1960's—and favorable price trends of the same magnitude, Brazilian recovery involved even more rapid import expansion. The current account balance moved from a surplus in 1965 and 1966 to a deficit of 2.3 percent of *GDP* in 1971–73; by the end of 1973 the external debt registered 17 percent of *GDP* compared to about 10 percent in 1967.

That rapid rise and the concomitant acceleration of product growth correspond to debt-led, rather than export-led growth. External resources, predominantly on commercial terms as Brazil for the first time became a factor in the Euro-dollar market, guaranteed the availability of foreign exchange for voracious import requirements. Those resources also permitted a surge in investment that did not have to be financed domestically at the expense of consumption. Finally, they rendered unnecessary the further

reform of the internal capital market that would have been necessary to achieve equivalent financial flows.

External indebtedness continued to expand after 1973 at even more rapid rates. The quadrupling of oil prices caught Brazil at a time when internal bottlenecks and cyclical excesses were creating internal economic adjustment problems and inflation was resurgent. Especially vulnerable to the rise in oil prices because 80 percent of its oil consumption needs were satisfied by imports, Brazil opted to postpone its adjustment to the external imbalance by relying even more heavily on debt. By the end of 1978 the external debt had mounted to more than $40 billion and 25 percent of *GDP*.

Debt-financed adjustment has produced less positive economic returns than the earlier phase of debt-led growth. Aggregate expansion has proceeded at rates almost half of the earlier period, and with a stop-go alternation that reflects the absence of a determined strategy to deal both with the external crisis and internal imbalances in agriculture, not to mention accumulating social disparities. Debt management has been poor, permitting massive acquisition of reserves and domestic monetary expansion. Limitation of imports—held virtually constant in nominal terms between 1974 and 1978 through stricter controls—has created difficulties. Inflation, at high levels since 1974, has accelerated recently to rates approaching 100 percent. Only in December 1979 has the government recognized the extent of the crisis and moved to counter it with a new liberalization package and a new strategy.

Even the economic miracle and its aftermath illustrates the theme: the external market affords opportunities, but also introduces constraints. Favorable and new access to capital markets that contributed to earlier growth has subsequently facilitated inadequate and inappropriate economic policies that now bind further options. Brazil, earlier appearances to the contrary, has still not definitively resolved the tensions of international market integration that have been present for over a century.

V

Reasonably reliable estimates of Brazilian gross domestic product suggest an average annual growth rate per capita of almost 3 percent per capita from the beginning of the twentieth century to the present.[2] The postwar period shows clear acceleration compared to what had gone before, with a rate of expansion in excess of 4 percent compared to 2 percent earlier. Such acceleration is tied to more rapid industrial growth: in per capita terms, agriculture shows much more modest increase.

These rates far exceed the performance of the core industrial countries over the long term. They clearly demonstrate that the current gap in income levels is a heritage of the more distant, rather than recent, past. Nineteenth-century estimates of Brazilian growth based on monetary and trade statistics confirm the likelihood of much more modest increases for much of the period. So does the simple backcast of national income figures. If even the prewar level of increase of product had prevailed over the last fifty years of the nineteenth century, the ratio of observed exports to assumed income would approach two-thirds—an implausible level well beyond that observed at the end of the century.

Brazil, like other tropical countries, therefore only began to experience the benefits of the spreading international economy quite late—despite involvement in world trade from colonial days. Backwardness did not produce an institutional response early enough to allow the forces of modernization and industrialization to translate into sustained advances in production. Trade and capital flows by themselves were relatively modest impulses to Brazilian expansion. In this respect some of the authors interpreting Brazilian development as inherently dependent exaggerate the role of the international

[2]These data are taken from Claudio Haddad, and subsequently from the official national accounts. Haddad's estimates from 1920 on are similar to those of myself and Villela and Suzigan; before 1920, they are based on more limited coverage. Synthetic estimates before 1900 based on trade and monetary data are more risky and should be used with caution.

economy. Indeed, one might argue that the nineteenth-century problem was preeminently one of too little, rather than too much, international integration.

Even later, as São Paulo coffee expansion first sparked more generalized economic development in Brazil, the share of exports to total product remained relatively small —certainly no more than one-quarter. That limited openness has frequently permitted Brazil greater autonomy and selectivity in dealing with the international economy. Pragmatic heterodoxy is a Brazilian hallmark. So too has been a continuing stress on industrialization. The relative success both have enjoyed is reflected in aggregate rates of expansion since the beginning of the century that few other developing economies can claim.

REFERENCES

A. Fishlow, "Origins and Consequences of Import Substitution in Brazil," in Luis Di Marco, ed., *International Economics and Development*, New York 1971.

C. Haddad, "Crescimento do Producto Real Brasileiro—1900/1947," *Rev. Brasileira Econ.*, 1975, *29*, 3–26.

N. Leff, "Long-Term Brazilian Economic Development," *J. Econ. Hist.*, Sept. 1969, *29*, 473–93.

______, "Economic Retardation in Nineteenth-Century Brazil," *Econ. Hist. Rev.*, Aug. 1972, *25*, 489–507.

______, "Tropical Trade and Development in the Nineteenth Century: The Brazilian Experience," *J. Polit. Econ.*, May/June 1973, *81*, 678–96.

Carlos Palaez and Wilson Suzigan, *História Monetária do Brazil*, Rio de Janeiro 1976.

Annibal B. Villela and Wilson Suzigan, *Politica do Governo e Crescimento da Economia Brasileira, 1889–1945*, Rio de Janeiro 1973.

DISCUSSION

MORRIS D. MORRIS, University of Washington: Two decades ago, David Landes wrote: "Western Europe...was already rich before the Industrial Revolution–rich by comparison with other parts of the world of that day and with the pre-industrial world of today." (See his book, *The Unbound Prometheus*, p. 13.) This judgment merely gave further currency to Simon Kuznets' more general statement that the pre-industrial per capita income level of developed countries was "several times" that of most underdeveloped countries today. (See his book, *Economic Change*.) The general conclusion is that today's less-developed countries (*LDC*) face a much more difficult development problem than faced nineteenth-century industrializers, *and* that this difficulty derives primarily from the sheer poverty of these countries as measured by per capita income. Recently, however, a flurry of essays have directly challenged this view.

The Heston-Summers essay points in a somewhat similar revisionist direction. Their results, with which I have no quarrel, should tend to dampen the rather extreme views that have derived from Kuznets' suggestion. The difference between India and *U.S.* per capita incomes in 1870, when adjusted by purchasing power parities (*PPP*), gets reduced to 4 or 5 to 1, a fairly modest ratio for a date when the United States was already a highly developed country and India was at the bare beginning of its very slow modern development. Pushed back to an even earlier era, the Indian–*U.S.* difference would certainly amount to very much less than that.

One important conclusion Heston and Summers draw is that a successful development effort in South Asia (and by implication in *LDC*s generally) will require many fewer resources than calculations based on official exchange rates suggest. Referring to the heterogeneity of South Asia, Heston and Summers remark that there are "several very large regions (each) with populations in the tens of millions, that have shown aggregate growth not so much lower than many industrial countries..." These regional "success stories" should affect our interpretations of Indian economic history.

The weaknesses of international income comparisons based on official exchange rates have been widely known. Heston and Summers now convincingly illustrate the even greater danger of making these comparisons over time. But policymakers and scholars feel free to ignore the implications of *PPP* adjustments. Similarly, I fear that having made a ritual obeisance to the Heston-Summers thesis about historical trends, development analysts will continue to operate as if this remains the best of all worlds.

Important a contribution though it is, the *PPP* adjustment still leaves many difficult problems unresolved. Purchasing power parity does not cope at all with the troublesome implications of differences in social organization among countries about which Kuznets wrote so long ago. Moreover, the Heston-Summers *PPP* adjustment cannot satisfy those fundamental conditions which are crucial if we are to think about the international welfare implications of these income comparisons. I refer here to questions about the stability of preferences and consumption bundles over time and about the meaningfulness of a social or national welfare concept.

Focusing on per capita income, even when the data incorporate the Heston and Summers improvement, is not the way to a satisfactory understanding of the development process. This is not only for reasons already advanced, but because economic historians are beginning to recognize the widely diverse patterns of development through which even the already developed countries have moved. The strategic analytic features lie not in the levels of per capita income attained but in the institutional arrangements which help or hamper the achievement of specific objectives.

The importance of this theme emerges most clearly when we examine real welfare

performance. This is something that cannot be captured by available per capita income measures but a recently developed Physical Quality of Life Index (*PQLI*) combining infant mortality, life expectancy at age one and literacy does reflect a quite wide range of social results. On a 0–100 scale, Guinea-Bissau with a per capita income of $120 was the poorest performer: *PQLI* = 12; Sweden with a per capita income of $7,668 was best: *PQLI* = 97. While there is some general relationship between per capita income and *PQLI* ranking, the correlation is surprisingly weak. (This appears to be true even when comparisons are with incomes converted at purchasing power parities.) For example, Saudi Arabia with a per capita income of $11,779 has a *PQLI* of only 31 while Sri Lanka (per capita income = $179) has a *PQLI* of 82. India (per capita income = $133) has a *PQLI* of only 43 but so does Iran even though it has a per capita income nine times larger than India's. In fact, one of India's poorer states, Kerala, had a *PQLI* of 70 in 1971.

What emerges from an examination of various poor country "success stories" (China, Cuba, Korea, Sri Lanka, Thailand) is that by design, force of circumstance or accident, these countries have generated institutional arrangements which enable them to buy very high life quality results—the ultimate objective of which we speak—very cheaply. What Heston and Summers tell us is that even if we measure performance in an orthodox way, the development resource burdens are not likely to be as great as has previously been feared. What the *PQLI* suggests, however, is that to achieve the basic human objectives of development requires more than rising per capita income. It requires specific kinds of institutional innovation and it is to this that we should turn our attention. This is not likely to be an easy task; it is most likely to be successfully done only when the effort is made by the people whose lives are involved.

SAMUEL A. MORLEY, Vanderbilt University: The subject of Albert Fishlow's paper is the extent to which the links between Brazil and the international economy generated or held back new development. This is an old and controversial subject among Brazilianists, and Fishlow himself has made several important contributions to the debate in the past. This paper is basically a summary of that earlier work. It is a nice summary, but adds little that is new.

Does integration of an economy into the international economy stimulate or retard development? In Brazil, two schools of thought have developed: the exports as an engine of growth view and the adverse shocks view. The former point to the beginning of Brazilian industrialization in São Paulo, site of the coffee boom in the late nineteenth century, the retardation of industrial growth during World War I, and the performance of the Brazilian economy after 1968 when exports led the economy to the highest rate of growth ever achieved over any five-year period between 1968 and 1973. The opposite point of view is based on the striking empirical fact that Brazilian industrialization and growth accelerated when the economy was forcibly cut off from world trade by the depression in the 1930's, by World War II and then in the 1950's by conscious government policy.

Fishlow gives some comfort to both schools of thought, asserting that industrial growth in the nineteenth century was not very significant, but also that industrialization under adverse shocks was probably not a very efficient development strategy because import constraints forced Brazil to adopt labor intensive, antiquated production techniques which made its consumer industries uncompetitive in the postwar period. However, his general position has long been that the outward orientation was less favorable to development and the import substitution policy more favorable than the advocates of an outward looking strategy maintain.

In support of this position Fishlow argues that nineteenth-century industrialization was insignificant and that export growth in São Paulo did not have much of a positive effect on the overall growth rate of the economy. To determine the extent to which exports were or were not a stimulus one should look at the experience of São Paulo, not Brazil as a whole. Clearly there was a

marked divergence in growth rates between the Northeast, which was unlucky enough to be specialized in the wrong export crops, cotton and sugar, and São Paulo where a coffee boom was going on. Exports generated income, some of which was invested in domestic manufactures. But the labor markets of the two areas were not sufficiently integrated for expansion in the South to absorb the excess labor of the Northeast. Labor was imported instead. Part of the increase in income in the South went into an expansion of the Brazilian population through immigration rather than an increase in per capita income. But that does not prove that exports were not "an engine of growth."

As a second argument in favor of the more inward-looking growth strategy, Fishlow argues that the *IS* growth of the 1950's created the industrial base which made possible the rapid industrial expansion during the "miracle" export boom. While I agree with that, this same agreement can be turned around. Brazil was able to surmount the adverse shock of the depression and begin a period of rapid sustained industrial growth in 1930 because she already had a rudimentary industrial base which made import-saving industrialization possible. Countries without such a structure were unable to respond so creatively to this shock. The consumer and intermediate industries that existed in 1930 mostly trace their origins to the outward-looking period of growth 1880–1914. In other words, one could argue that the export boom made possible Brazil's subsequent successful import substitution. It is probably safer to say that both orientations have made contributions to growth at different periods.

Most writers have thought of the "miracle" period as export led growth. Fishlow proposes a new label: "debt-led growth." This strikes me as misleading. Rapid export expansion was a demand stimulus while increasing imports were either an offset to that stimulus in the case of import substitutes, or allowed an expansion in supply. Which sort of imports dominate isn't clear, but in any case, the balance-of-payments gap wasn't significant until 1972. Brazil's improved credit rating allowed it to borrow to increase reserves—not just imports. Net debt increased from $3.5 billion in 1968 to $6 billion in 1973, not a drastic increase considering that exports more than doubled during the same period.

Rather than focusing on external debt in the miracle period, Fishlow should have focused on excess capacity. Because of five years of orthodox stabilization in the mid-1960's, Brazil had substantial excess capacity in 1968. This meant that growth required surprisingly low rates of capital formation—no more than 20 percent for a growth rate of 10 percent prior to 1972. Supply expansion was therefore not a problem as Brazil moved to take advantage of the exceptionally favorable export markets created by the Vietnam War boom. However after 1972 the investment ratio rose to over 24 percent. Brazil's balance-of-payment problems were caused as much by high capital goods imports as they were by the high price of oil.

Thinking further about the Brazilian experience, I am struck by the key role that demand has played relative to supply, which is just another way of saying that in the past supply expansion was not a constraint. This is a rather Keynesian viewpoint, but it appears that rapid growth almost invariably followed any significant expansion in domestic demand. At times exports have been the source of that demand as for example in the late nineteenth century or the recent "miracle" period. At other times adverse shocks and supportive government policy have been the source. In the 1930's a severe reduction of import capacity coupled with governmental maintenance of internal income through purchases of coffee guaranteed a buoyant market for domestic production. In the postwar period the government engaged in a conscious policy of import substitution in key sectors at the same time that it increased deficit spending. In the "miracle," a reversal of orthodox stabilization with easy credit and heavy subsidies to export activities guaranteed a dynamic growth in demand. In all of these cases growth occurred because the Brazilian economy responded with capital formation

and output growth to profit signals. The result has been seventy-five years of very impressive economic growth.

But in the future the supply constraints are going to be much more difficult than they have been in the past. As we have seen from investment, the capital output ratio must have risen substantially. So has the import coefficient. This means that growth is going to require both a greater saving and export effort, both of which are going to be hard to achieve. The years ahead will be difficult ones where supply constraints and planning to overcome them will be far more important than the simple provision of profit incentives through demand creation were in the past. The optimistic lessons of Brazil's past history may well be misleading signposts for the future.

THE RECOIL FROM WELFARE CAPITALISM: POLITICAL AND SOCIOLOGICAL PERSPECTIVES

The Welfare State in Trouble: Systemic Crisis or Growing Pains?

By ALBERT O. HIRSCHMAN*

We all know about persons who are *complacent*: they gloss over an important flaw in the functioning of something—a human body, a marriage, an economic policy, or a society—and try hard to convince themselves and others that nothing is really wrong; if they advocate any action, once the symptoms of trouble can no longer be ignored, they will typically prescribe aspirin when radical surgery is required. What about a term for the opposite fault: a term, that is, that would designate a person who forever diagnoses fundamental disorder and prescribes radical cures when the difficulty at hand may well take care of itself in time or only calls for mild intervention? I propose, for want of a more compact term, the "structuralist (or fundamentalist) fallacy," since those who are affected by this trait always speak of structural problems and the need for fundamental remedies.

Economists have long arrayed themselves into the two camps that are implicit in the two opposite faults just noted: those who are convinced that every departure from equilibrium is likely to be temporary and requires only a bit of clever management (if anything) have long been battling it out with those who are just as certain that such a departure signals a deep malady and perhaps the final crisis of the system. Naturally enough, the former turn out, from time to time, to be complacent while the latter will on occasion be found to commit the structuralist fallacy. Being rather bored by both ideological camps, I tend to shift from one to the other depending on whom I am talking to. Given the complexity and ambiguity of the real world, a useful function may actually be served by such contrary behavior as I now hope to show by looking at the so-called "crisis of the welfare state."

That the welfare state is in trouble can hardly be contested. Considerable difficulties are presently experienced in the West in extending or even maintaining the social accomplishments of recent decades. Hostility toward some of the services provided is widespread, even among the beneficiaries. A political reaction has already occurred in some of the countries (Sweden, United Kingdom) that have gone farthest along the welfare state road. In the United States, Proposition 13 and assorted phenomena have generally come under the label of "tax revolt"; to an important extent, this revolt is not a purely self-serving act on the part of the taxpayer but arises from a growing lack of confidence in the state's abilities to "solve" social problems. This lack of confidence can in turn be traced to recent experiences in social engineering (for example, the War on Poverty) and to their alleged failure.

A number of explanations are being offered for the difficulties the welfare state is running into and most of them are of the structuralist kind. Long ago, Colin Clark alleged that in the nature of the capitalist system a fairly rigid limit was set to the ability to divert factor income for purposes of expanding social services and other public expenditures: the system would stop working (that is, capitalists would no longer invest, workers' productivity would fall off, and so on) if that level were exceeded. The inability of the system to stand more than a certain level of transfer payments was

*Professor of social science, Institute for Advanced Study, Princeton.

thereby declared to be one of its structural properties. Widely espoused at first by conservatives, this structuralist thesis was given a Marxist tinge in James O'Connor's book *The Fiscal Crisis of the State* which explained this crisis in terms of the underlying clash between two basic functions of the capitalist state, that is, between the need to assure continuing capital accumulation, on the one hand, and the imperatives of legitimation through some redistribution of income, on the other.

Subsequently, Jürgen Habermas worked the O'Connor thesis into a more general argument in his book *Legitimationsprobleme im Spätkapitalismus* (published in English as *Legitimation Crisis*). From there the argument passed back to the conservative or neoconservative camp which stirred up, in the mid-1970's, a considerable debate around issues of "governability of the democracies" and "governmental overload." Once again, the absolute and relative expansion of public expenditures for health, education and welfare, termed the "welfare shift" by Samuel Huntington, was viewed as an important ingredient of a widely proclaimed "crisis of democracy" (see Michel Crozier et al.).

A very different explanation why the welfare state is in disrepute if not in crisis is implicit in the argument of the late Fred Hirsch in *The Social Limits to Growth.* According to this work, the difficulty about increasing social spending beyond a certain point does not originate only among those who are made to pay for it; to an important extent, it arises from the unhappiness of those whom the spending was supposed to benefit. The reasoning is based on the concept of "positional goods" which are desired less for their own sake than for the social rank and distinction they are expected to confer. Recently governments have been getting into the business of supplying such goods on a mass basis. In education, for example, certain diplomas have become more widely available, but for that reason their possession no longer leads to a better paying job or higher social position. As a result, the recipients of the publicly provided services are likely to be disgruntled and cannot be counted on to support the expansion of public spending and social services.

Once again we have here an argument of a structural kind. It is now the frustration of its very clients that explains the crisis of the welfare state whereas the earlier conservative and neo-Marxist theories stressed the adverse reactions of the suppliers of investment capital. Both arguments see the expansion of the welfare state as creating a contradiction with basic characteristics of economy and society: the ensuing crisis is of a "fundamental" nature and requires "radical" remedies if it can be solved at all.

I shall now suggest an alternative explanation. It can be argued that a rapid expansion of the supply of certain goods and services is likely to bring with it a *deterioration in their quality* in relation to expectations and that it is this quality decline which produces disaffection with the performance of the public sector. If this argument has merit, the problem is not all that fundamental, for the quality decline may well be temporary.

The idea that an expansion in output will be detrimental to product quality is a priori plausible, but has not, as far as I know, been subjected to a great deal of economic analysis, in contrast to the enormous attention that has been lavished on the effect of increasing output on unit cost. It is the old story of the neglect of the qualitative. An obviously interesting question is: when will an expansion of output caused by increasing demand be accompanied by a significant decline in quality? In line with this question one might define an *output elasticity of quality* which would measure the response of quality to an increase in output within some stated time interval. Like the price elasticity of demand, this elasticity would normally have a negative sign with a limit of zero for those goods whose quality is totally unaffected by output increase. The output elasticity of quality is likely to be strongly related to the substitutability of inputs. If all input coefficients are rigidly fixed, an identical article, be it an umbrella or an airplane, is always being produced and quality cannot deteriorate as output is expanded. The possibility of deterioration opens up when one input or factor can be substituted for

another. The smooth isoquant of the textbook is of course drawn on the assumption that an increase in the use of one factor makes up *totally*, insofar as quality of the product is concerned, for the decreased use (in relative terms) of another. Actually it happens all the time that some factor or input substitution that is carried out to satisfy an expanding demand (one factor or input being in inelastic supply) results in a product that is not up to the traditional standards. Factor and input substitutability thus open the door to quality deterioration whereas fixed coefficients make for invariant quality. Since substitutability is the rule in the world of neoclassical economics it may be surprising that changes in the quality of output have not been given more attention. The reason is another pervasive assumption: that is, perfect information or instant learning in a competitive market on the part of consumers who immediately adjust their demand upon being faced with an inferior product and enforce an appropriate relative price reduction for it.

In the real world, of course, input shifts and the resulting quality deterioration are often combined with noncompetitive markets, consumer ignorance and slow learning about the changed characteristics of the product. It is my contention that this combination of circumstances is precisely characteristic of certain social services whose expansion in response to widespread demands has been considerable in recent decades.

Education is a pertinent example. On the production side, this "article" has a particularly high tolerance for quality decline and low-level performance as expanded educational services can be and often are offered in spite of various unresolved bottlenecks, that is, with unprepared teachers, impossibly crowded classrooms, inadequate library and laboratory facilities, and so on. Educational services are in fact an extreme illustration of the possibility of quality decline which originates in a lopsided increase in inputs: it would be quite unthinkable to market similarly defective refrigerators or airplane services.

The latitude for quality decline is also related to the demand or consumption side. Consumers are poorly informed about the quality of the expanded services of an educational system, and have few alternatives to choose from. Nevertheless, if the newly offered services are in fact defective the result will eventually be widespread disappointment and discontent. The damage inflicted by having received a poor education is not easily undone: unlike apples, education is not bought recurrently in small quantities. This may lead to disappointment which cannot be easily extinguished by doing the right thing next time in the market and therefore will lead a life of its own—with some possibly serious social and political consequences.

The welfare state may thus face a wave of hostile public opinion and as a result may well pass through a difficult phase, with the need for consolidation and even retrenchment. However, if the reasons suggested here for the change in the climate of opinion are correct, the trouble could be temporary. The loss of popularity is connected with the decline in quality which in turn is due to temporary factors, such as the rapid increase in supply. Once the new services can count on an adequately expanded base of inputs, the quality decline, which is here seen as the basis for public disenchantment with the welfare state, should be duly reversed and the social advances consolidated in fact as well as with public opinion. Over the long term, in other words, our output elasticity of quality is likely to be closer to zero than over the short term.

This "nonfundamental" diagnosis of the present difficulties of the welfare state is reinforced by some further observations. In the first place when certain social services like education are expanded so as to cater to newly emerging groups, it may not be appropriate to offer exactly the same services as have previously been supplied to the traditional "educated class." Hence even without quality decline, and precisely because there has been no change or adaptation, the services might be ineffective and meet with criticism and resistance. Again a period of learning and mutual adjustment will be needed.

Secondly, there has been a tendency in recent years for the demand for certain

services to arise in advance of real knowledge how to satisfy it; examples are day-care facilities and psychotherapeutic services. What happens in these situations is that suppliers only begin to learn on the job, in the process of rendering these newly popular services as best they can. A great deal has been written by Kenneth Arrow, George Akerlof and others about consumer ignorance and the resulting asymmetrical situation of consumer and producer. In the present case, producers and suppliers of services are just as ignorant as consumers, at least during the earlier stages of their operation, and this accounts for the poor quality of services rendered and for the consequent disappointment of the consumer. Here also, a learning process will take place, eventually leading to better-quality service and to more correct consumer expectations.

In sum, the difficulties of the welfare state can be interpreted, in part at least, as growing pains rather than as signs of systemic crisis.

To conclude I wish to raise a question in the sociology of knowledge: why have the various conceivable nonstructural arguments not been coherently put forward so far, with the result that we could only choose between various kinds of structuralist explanations? The reason, I think, lies in a rather odd ideological asymmetry. In explaining it briefly I return to my introductory remarks on structuralist vs. nonstructuralist approaches to problem solving. Structuralist thinking about a problem or crisis comes easily to those who dislike the institution that experiences the problem or finds itself in crisis. For example, right-wing and conservative people dislike the welfare state and oppose its expansion: they are naturally prone to interpret any difficulties it encounters as symptoms of a deep-seated malady and as signals that radical retrenchment is in order. For similar reasons, left-wing and liberal opinion has traditionallly opted for structuralist explanations when it came to account for difficulties experienced by capitalism. But with the debate about capitalism and the market economy having stood in the center of public discussion for so long, this tradition appears to have created on the Left something of an unthinking *structuralist reflex*: Left-liberal people are automatically partial to structuralist explanations, even though ideological self-interest ought to make them diagnose some difficulties—those that affect structures they themselves have promoted—as self-correcting or temporary. As a result of this strange ideological trap into which the Left has been falling, there has been a marked lack of balance in the analysis of the current difficulties of the welfare state.

REFERENCES

Michel Crozier et al., *Crisis of Democracy*, New York 1975.

Jürgen Habermas, *Legitimationsprobleme im Spätkapitalismus*, trans. as *Legitimation Crisis*, Boston 1975.

Fred Hirsch, *The Social Limits of Growth*, New York 1976.

James O'Connor, *The Fiscal Crisis of the State*, Cambridge, Mass. 1973.

Accounts of the Welfare State and the New Mood

By IRA KATZNELSON*

The radical shifts in electoral fortunes and public philosophy that have called into question the postwar relationship of capitalism, democracy, and the welfare state are symptoms that invite diagnosis. Alternative assessments depend, of course, on more general theoretical understandings of the welfare state—its genesis, development, and consequences. The quality of diagnoses of the current hard times of the welfare state cannot be independent of these more general views. As Theodor Geiger observed in *Die Masse* and *Ihre Aktion,* "...counterrevolution is part of revolution. Not that it is a revolution. Rather counterrevolution belongs to it, results from it in the form of a refluent movement. It is literally a *reaction*" (p. 59).[1] I take this observation as my point of departure. My main purpose is to suggest an alternative to available views on the growth of the welfare state, and to argue that current events may best be understood as integral to the story of its development.

Treatments of the welfare state have been concerned above all with two questions: What accounts for the rather spectacular growth in this century of welfare state expenditures? What are the effects of the welfare state on the persistence of market capitalism? Answers to these questions have produced four clusters of explanation. The first treats the main sources of welfare state growth as political. Party and group competition in mass franchise democratic regimes is the motor of expansion, and the welfare state is an adjunct to the market. The second also sees the political process as the key source of state growth, but understands these developments as a threat to the dominance of market principles. The third view disagrees on both counts. It stresses the links between capitalist and welfare state growth, understanding the latter as a necessary precondition for private capital accumulation (see Ian Gough). The economy not only tolerates but requires an increasingly interventionist state. The last approach agrees with the third on the sources of state growth but not about its consequences. Like the second it stresses that as the welfare state develops the nonmarket sphere of society is enlarged, and that the discipline and even the rationale of market relations is undermined. The welfare state is subversive of capitalism.

These familiar families of explanation share many traits. Their causal statements have a functional logic. What is explained —the development of the welfare state—is determined by its effect on what explains it; that is, on democratic processes or on the capitalist economy. Democracy requires results that modify market allocations if it is to induce mass participation. Capitalism requires social welfare policies to make its own reproduction possible. The advocates of these positions ask us to select between them, yet the grounds for making this choice are not apparent. The selection of capitalism or democracy as the factor to be functionally paired with the welfare state is arbitrary, since the development of welfare policy coincides with the emergence of mass franchise democracies and with "second industrial revolution" capitalism. The customary explanations also ask us to make a second choice between supportive and baleful effects on the capitalist economy. But this choice too seems rather forced, ignoring contradictory possibilities. Further, the traditional views provide plausible (and not easily refuted) claims about the growth of domestic welfare state activities in all the capitalist democracies. But they have nothing to say about the variations in social policy that differentiate these societies. Indeed, without elaboration of the mechanisms that sustain the functional relation-

*University of Chicago.

[1]Cited in Arno Mayer's *Dynamics of Counterrevolution in Europe...,* pp. 44–45.

ships between democracy/capitalism and the welfare state, an understanding of specific national situations in comparative perspective is not possible. Such an elaboration may be of an intentional or nonpurposive kind. Most treatments of the welfare state either fail to provide such an elaboration, or choose arbitrarily between each of the two possible kinds; stressing a single favored mechanism. Thus, the political perspective may stress the leadership of key civil servants, the ratchet effect of the outcomes of prior conflicts, or the disjunction between concentrated benefits and diffuse costs. The economic perspective may emphasize self-conscious demand management to deal with problems of underconsumption, the prisoners-dilemma situation of competing capitals and the need for a "supra capital," or the supports required for smoothly operating labor markets. The warrants for these choices as the main mechanisms sustaining the functional relationship are usually weak at best.

An alternative to the limits and choices imposed by the inherited major views on the welfare state would have to meet the test of accounting for the double correlation of the welfare state and advanced capitalism, on the one hand, and mass democracy, on the other. Welfare, as Raymond Williams has reminded us, is an old term, used at least since the fourteenth century to refer to prosperity or well-being. But its sense as the object of the organized provision of care or services is only seventy-five-years old. "Welfare policy" he dates from 1905; the "welfare state" only from 1939. The modernity of these terms underscores the importance of both correlations which have been understood in functional terms. An alternative would also have to be able to come to terms with the contradictory character of the welfare state, and of social democratic movements and programs. Social democratic programs have been adopted everywhere in the West (even in countries where social democratic movements are weak or nonexistent), and it is impossible to imagine that any capitalist society may achieve stability and continuity without a dopting welfare state policies. But we do not possess a theory that tells us precisely what the requirements are for a capitalist society. Specific policies may be adopted even when not required by capital, and their premises and operation may indeed make capitalist reproduction more difficult. Moreover, an alternative would have to be capable of making sense of the distinctive configurations of social policies in different capitalist democracies, and thus make sense of the "league tables" of welfare state expenditures. Finally, such an alternative, if functionalist in its first order of explanation, would have to provide the second order elaborations missing in the traditional accounts. Before sketching what such an alternative might look like, a few remarks about the meaning of recent developments are in order.

That there has been a vigorous assault on the legitimacy and practices of the welfare state (understood here to be government services, social insurance, and other transfers that reshape and compensate for allocations of the market) is not in doubt, even though the boundaries of this challenge are far from clear. This assault is part of a larger skepticism about the growth of government and the state's ability to solve social problems. In a more focused way, it is also part of a reappraisal of the social democratic agenda more generally, which includes not only welfare state programs but the selective nationalization of industry, mechanisms of economic and social planning to incorporate nonmarket priorities into economic and spatial development, and the use of macro-economic policy tools to minimize unemployment rather than inflation and to promote a redistribution of resources.

The meaning of these symptoms is unclear both in relationship to the past and the future. Just yesterday, the majority of observers talked about the welfare state in terms of uninterrupted growth. Plaints about this trend and proposals to undo it seemed utopian and quixotic. Political analysis sought to account for this policy direction and to assess its consequences for stability and change. Seen against this background current events may appear to prom-

ise a very different future trend. At a minimum, I am doubly agnostic. It is far from clear that the welfare state is in greater difficulty today than, say, a quarter-century, or half-century, ago. The chronological comparisons that provide justification for considerations of the recoil from welfare capitalism have a rather more condensed time frame. Further, it would be folly to extrapolate long-term trends from the current mood, since there is precious little evidence that antistate rhetoric is being matched by actual rollbacks of social democratic achievements. I am reminded of the story my colleague Adam Przeworski tells of the empiricist who jumped from the Empire State Building yelling "so far so good" as he passed each floor.

In spite of these cautions, the symptoms of antigovernment assertiveness do demand diagnosis. I am less convinced that the tools of diagnosis must be crafted exclusively to account for the 1970's. Quite the contrary. Our capacity to come to terms with recent events hinges on whether we have a more general approach to welfare state development. Most explanations of the current "crisis" of the welfare state do in fact depend on the four conventional families of views.

These accounts usually fall in the second and fourth categories. The neoconservative variant argues that a surfeit of democracy has pushed welfare state expenditures beyond the level the economy can stand without paying high productivity and investment costs. In the Marxist variant, the displacement of contradictions of production and distribution to the state by enlarging the nonmarket sphere has produced an evermore precarious balance between capital and the state, and a challenge to the ability of either to deliver expected goods and services. I think it accurate to say that in the 1970's there has been a parallel shift in the locus of conventional accounts from the first to the second cluster of views by those who see democratic politics as the main cause of welfare state development, and from the third to the fourth category by writers on the Left who are rather less impressed than they once were with the reproductive capacity of welfare capitalism. The movement from the first to the second family has been the dominant intellectual fashion in recent years, and it has produced a host of suggested solutions that aim to place restraints on democracy in the interest of governability and economic rejuvenation. By contrast, the movement from the third to the fourth sets of views has taken on more and more the character of abstract and academic debate about possibilities of reformist transitions to socialism at the moment when more and more of the constituents of Left parties have been convinced by the ideologically dominant second point of view. This shift has been understood in various ways. Workers may be discovering that they have received what benefits they could from the welfare state and now must pay too many of the costs. Or, as Albert Hirschman suggests in his essay for this session, the quality of services may disappoint consumers in a period of rapid expansion and an inadequate base of inputs. I take it that such views are meant to explain why specific intellectual and ideological constructs may have mass appeal at a given moment, and for this purpose have some force, but they do not address the quality of the conventional approaches to the welfare state.

I do not have a fully developed alternative to propose, but I should like to suggest that the sources of welfare state growth need not be exclusively "political" or "economic" and that the character of social policy may support the recreation of the capitalist order while being at least potentially antagonistic to it.

The state's social policy activities may be said to develop in two interrelated ways. First, the ordinary operation of the capitalist political economy requires a variety of state activities for the recreation of capitalist productive and social relations. It seems beyond doubt that a reproduction *minimum* of social policies is required to insure the accumulation process and to give it broad social acceptance. But the level and content of this state intervention in the market cannot be the subject of theoretically informed calculations which hold equally well for all the capitalist democracies. The

minimum varies from state to state and from time to time because it represents an amalgam of what has come to be economically, politically, and culturally necessary. Although largely the result of past class and group struggles, the minimum at a given moment is accepted by all classes in the society.

An institution "in being" defines reality for its participants, and constitutes a definition of problems and ways to work them out. The welfare state at the level of the minimum in a given society presents a common set of problems which appear technically and politically neutral for politicians, administrators, and client-citizens. Civil servants and contestants for the vote are compelled to respond to such problems. In constructing their responses, they may draw on elaborate and sophisticated tools from welfare economics and theories of public finance which provide a framework for the identification of an appropriate welfare state minimum. This technical literature understands state growth as the product of social overhead requirements and external costs generated by private sector development, and it seeks to find alternative guidelines to public sector expansion in the absence of price tags linking specific outlays to specific contributions by taxpayers.

But the welfare state does not only develop in this "problem-solving" way. Under certain conditions various political parties and social movements may succeed in utilizing the democratic process (including protests as well as elections) to push welfare state advances forward at a pace more rapid than that dictated by the emergence of manifest "problems" within the consensual framework of the reproduction minimum. When such attempts succeed it makes sense to speak of a social policy *surplus*, required neither by the logic of capitalist nor democratic relations (understood independently from the tension between the functional requirements of capitalism and democracy and the dynamics of volitional class and group organization and demands in specific settings).

The level of welfare state expenditures as well as their character at a given moment is the sum of the reproduction minimum about which there is broad communal consensus, and the results of current, factional, and class conflicts. The dimensions of the minimum inherited from the past, and struggles about whether and to what extent a surplus will be fashioned in the present, are functions of the capacities of organized groups and classes. For this reason, interpretations of differences between the welfare state policies of different countries requires not only a movement away from simple first-order functional explanations that fit all the cases, but also away from treatments of class and class struggle defined in unmediated fashion by the position of classes in the mode of production. Much as state policies vary from one capitalist society to another, so do the character of groups and classes sharing dispositions and capable of organization and action.

I should stress that it does not follow from these considerations that policies now part of the "minimum" are necessarily more functional for market capitalism than those generated by struggles about the "surplus." In part this is the case because as G. A. Cohen argues, "class insurgency is more likely to achieve its object when the object has functional value...There is victory [for the working classes] when capitalism is able to sustain itself only under the modification the reform imposes on it" (pp. 295–96). It is also the case because, in contradictory fashion, any expansion of the nonmarket sphere, at whatever level of welfare state development, entails at least short-term costs for investment and productivity, and for some privileged economic actors. Further, it should be recalled that a present minimum is yesterday's surplus, and, as such, was the occasion for struggle. For this reason it is difficult to say that the welfare state is in more trouble today than at many other periods. Struggles about the welfare state are structurally conditioned by the distinction between minimum and surplus. Struggles about where the line between minimum and surplus is to be drawn, and about the potential surplus itself are endemic to the welfare state. These struggles consist in a set of limits which are set in part by the functional

requirements of capitalism and democracy, in part by the inheritance of past policies and compromises, and in part by factors that affect class and group capacity.

I think these observations are of some use in diagnosing current symptoms. The logic of the welfare state that I have sketched has not changed. The new resolve of the Right and the incapacity of the Left to define the terms of political discourse have not altered the existing welfare state minima in the different societies of Western Europe and North America. Indeed, it is rather commonplace to note that once social policies cease to be targets of group and class contests, and once they provide a set of givens that intertwine other social and economic policies in complicated ways, they prove virtually impossible to undo. Conservative Republicans in power do not eliminate the Medicare and Medicaid programs they once fought so hard to prevent; Thatcherite Conservatives do not dismantle the British Health Service to return to fee for service medicine. The costs to the social order simply would be too high. Indeed, on the record, it is reasonable to say that governments of the Right have been compelled to address social problems in ways that actually expand the scope of the welfare state. What has occurred in the past few years has been an intensification of the resistance to this process of expansion of the "minimum," and concurrently, an erosion of the capacity of various social democratic parties and movements to secure a welfare state surplus.

The basic contours of the current situation, in short, are not new, but the balance of capabilities has altered. The causes (here I agree with Hirschman) are not to be found in the structural logic of the welfare state itself, nor principally (and here I differ with him) in the relationship of service bureaucracies to their clients. The most important reason has been the persistent stagflation of the post-1973 capitalist economies, which has had three main consequences for the welfare state understood in the terms I have suggested. First, stagflation has raised the perceived costs of welfare state expenditures for investment and productivity. Although even a cursory examination of aggregate growth rates in the postwar period contradict the argument found almost daily on the pages of the *Wall Street Journal* that there is an inverse relationship between social policy expenditures and economic growth (the *Journal* prefers to remember the British as it forgets the Scandinavians or Austrians), the economic crisis of the 1970's has made it much easier for the Right to argue in particular settings that the economic system would work better if only the share of national income spent on social services were to decrease. Second, social democratic parties were in power for significant periods in the early and middle 1970's, and they could hardly escape electoral recriminations. The Swedish SAP and the Labour Party in England were especially vulnerable to the charge that their management of the welfare state contributed to growing economic difficulties. Third, the inability of Keynesian nostrums to deal with the stagflation of the 1970's Left social democratic parties in and out of power (including the Euro-communists) without a coherent economic program with which to confront the promarket advocacy of the Right. The marriage between a program of Keynesian demand management and attempts to secure modest welfare state surpluses that had sufficed for the Left since the 1930's split up. The result was a series of unappealing choices. Left parties could either, as the British Labour Party did in 1976, announce that they were prepared to "roll back" the welfare state and claim that they could do a better job of restricting government than political parties to the right, or they could push ahead with the advocacy of programs for a larger welfare state (and, more generally, social democratic) surplus in the absence of a coherent stragegy to manage the existing capitalist economy. Such programs to expand the sphere of nonmarket rationality, including further measures of nationalization of industry, land, and credit have, of course, made their appearance. But these options are very difficult to convince voters to bet on because they invite investment rebellions by capitalists who wish to avoid allocations which are suboptimal with re-

spect to profits.[2] Proposals which might deepen the current economic crisis in the interest of future prosperity and equality hardly constitute a realistic basis for achieving electoral majorities. As a consequence, the Left has found itself wriggling once again on the head of a very old pin.

[2]For an excellent discussion, see Przeworski.

REFERENCES

G. A. Cohen, *Karl Marx's Theory of History: A Defense,* Princeton 1978.

Theodor Geiger, *Die Masse und Ihre Aktion,* Stuttgart 1926.

I. Gough, "State Expenditure in Advanced Capitalism," *New Left Rev.,* No. 95, July-Aug. 1975.

Arno J. Mayer, *Dynamics of Counterrevolution in Europe, 1870–1956: An Analytic Framework,* New York 1971.

A. Przeworski, "Social Democracy as a Historical Phenomenon," *New Left Rev.,* forthcoming.

Raymond Williams, *Keywords: A Vocabulary of Culture and Society,* New York 1976.

Equality, Incentives, and Economic Policy

By RANSFORD W. PALMER*

The future reduction of income inequality between whites and blacks in the United States is affected by the interaction of federal educational policy to encourage human capital investment and federal economic policy to achieve full employment and stable prices. The effect of a full-employment economy on income disparity between blacks and whites is limited by the occupational distribution of the black labor force. However, human capital investment to upgrade the occupational distribution of the black labor force is more effective within a full-employment setting which assures a higher rate of return on investment than in a setting of persistent high unemployment. Thus from the point of view of blacks, the scenario for the best of all possible worlds is one in which economic policy strives actively to achieve full employment while educational policy stimulates the growth of human capital investment to upgrade the occupational distribution of the black labor force. No one policy without the other can be fully successful in reducing persistent income disparity.

There is a general concensus that a full-employment economy is particularly beneficial for blacks, since a tighter labor market will increase black income and reduce black unemployment rates by a greater proportion than that of whites. The net result is a reduction of black poverty and some improvement in the distribution of income. The relationship between poverty and unemployment in the 1960's is particularly striking in Table 1.

Between 1964 and 1969 when black unemployment rates fell from 9.6 to 6.4 percent, the percentage of black families in poverty fell sharply from 40 percent to approximately 28 percent. This percentage appears to be some kind of a floor below which it has rarely fallen throughout the 1970–77 period, as the unemployment rate grew to over 13 percent. Needless to say, a proliferation of public welfare programs and the initial impact of the civil rights laws also played an important role in reducing poverty in the 1960's. In fact, it may be argued that expanded welfare programs combined with changes in the personal income tax structure helped to offset some of the negative income effects of rising black unemployment in the 1970's.

The difference in the behavior of the income disparity between blacks and whites (as measured by the ratio of the median family incomes of the two groups) in the 1960's and the 1970's is also readily observed in Table 1. Declining unemployment rates for blacks in the 1960's were accompanied by an identifiable improvement in the black-white median family income ratio, from .54 in 1964 to .61 in 1969. But in the 1970's, the improvement did not continue; instead, there appears to be a general hardening of the median income gap into a permanence around a ratio of approximately .60.

The proximate cause of this persistent income disparity between blacks and whites and among blacks is reflected in the dominance of low-productivity workers in the occupational distribution of the black labor force. In 1974, for example, the share of service workers in the total employed black labor force was 26.2 percent, compared to 11.8 percent for the employed white labor

*Professor of economics, Howard University. I wish to thank Surinder Gujral, Reynold Madoo, and Margaret Simms for helpful comments on an earlier draft of this paper.

TABLE 1—PERCENTAGE OF BLACK FAMILIES IN POVERTY, BLACK UNEMPLOYMENT RATES, AND THE RATIO OF BLACK TO WHITE MEDIAN FAMILY INCOME, 1964–77

Year	Percent of Black Families in Poverty	Black Unemployment Rates[a]	Ratio of Black to White Median Family Income
1964	40.0[b]	9.6	.54
1965	39.7[b]	8.1	.54
1966	33.9[b]	7.3	.58
1967	33.9	7.4	.59
1968	29.4	6.7	.60
1969	27.9	6.4	.61
1970	29.5	8.2	.61
1971	28.8	9.9	.60
1972	29.0	10.0	.59
1973	28.1	8.9	.58
1974	26.9	9.9	.60
1975	27.1	13.9	.62
1976	27.9	13.1	.59
1977	28.2	13.1	.57

Source: Bureau of the Census (1975; 1979), and Council of Economics Advisors.
[a]For blacks and other races. The unemployment rates for blacks only for 1975, 1976, and 1977 were 14.7, 13.8, and 13.9 percent, respectively.
[b]Includes other nonwhite races.

force. While the ratio of white-collar workers to service workers was 1.08:1 for blacks, for whites it was 4.2:1. The figures for 1977 registered only modest changes toward improving the ratio for blacks. One result of this occupational distribution is that income distribution among blacks is more unequal than income distribution among whites.[1] A measure of this inequality over the period 1947–74 is presented in Table 2. While the share of total black income received by the highest fifth of the black population declined marginally from 45.3 percent in 1947 to 44.2 percent in 1974, that of the lowest fifth increased only slightly from 4.3 percent to 4.6 percent. For whites, the share received by the highest fifth declined from 42.5 to 40.5 percent, while that of the lowest fifth rose from 5.5 to 5.8 percent. Over the same period, the ratio of mean black family income to that of whites grew from .54 to .68, suggesting that the reduction of income disparity between black and white families was not accompanied by any significant reduction in income disparity among blacks.

Casual empiricism indicates that, in general, the distribution of investment in human capital is influenced by the size of families and the distribution of family incomes. In 1974, black families with incomes under $5,000 had roughly 3.5 times as many family members between 18- and 24-years of age as black families with incomes over $15,000. Table 3 shows that those families with incomes over $15,000 sent 42 percent of their 18- to 24-year-olds to college, while those with incomes under $5,000 sent only 17 percent. In contrast, whites with incomes over $15,000 have five times as many 18- to 24-year-old family members as whites with incomes under $5,000, and sent 50 percent of their 18- to 24-year-olds to college compared to 17 percent for those white families with incomes under $5,000 (see Bureau of the Census, 1975). If the distribution of the propensities to enroll in college and the distribution of college-age children by family income remain as they are shown in Table

[1]Because much of the data on black income that are available in the government publications cited are included in the data for "Blacks and Other Races," a precise black-white comparison is not always possible. However, we assume that the data for "Blacks and Other Races" in Table 2 provide a fair representation of the black economic condition.

TABLE 2—DISTRIBUTION OF INCOME AMONG FAMILIES BY RACE: 1947; 1964; 1974

Year	Number of Families[a]	Lowest Fifth	Second Fifth	Third Fifth	Fourth Fifth	Highest Fifth	Top 5 Percent	Mean Income ($)
				Blacks and Other Races				
1947	3,117	4.3	10.4	16.1	23.8	45.3	16.4	1,986
1964	4,754	4.4	10.5	16.2	24.2	44.7	16.9	4,726
1974	6,262	4.6	10.0	16.2	25.0	44.2	15.9	10,192
				Whites				
1947	34,120	5.5	12.2	17.0	22.9	42.5	17.4	3,687
1964	43,081	5.5	12.4	17.8	23.8	40.5	15.7	7,625
1974	49,451	5.8	12.3	17.6	23.8	40.5	15.1	15,047

Source: Bureau of the Census, 1979.
[a]Shown in thousands.

3, the continuation of wide income disparities among blacks and between blacks and whites is assured. Indeed, the income disparity among blacks may even become sharper as the relatively few children of the relatively few families at the top of the black income pyramid inherit most of the black-owned property.

If a reduction of income inequality between blacks and whites is to result in an improvement in the relative income status of most blacks, it must be accompanied by a reduction of income disparity among blacks. Under traditional black-white income comparisons, it is possible that the reduction of the black-white income differential may be due largely to the acceleration of the incomes of those at the top of the black-income pyramid. That is to say, growth of aggregate black income may take place with no significant improvement in the distribution of income among blacks. The upward shift in the occupational distribution of the black labor force required to improve both the inter- and intraracial distribution of income can be accomplished only if low-income families make greater human capital investment in their children. But this is not possible because low-income families are caught in a vicious circle: they save at a lower rate than middle- and upper-income families, inherit no property, accumulate little capital, and therefore experience a rate of growth in their income that perpetuates their low-income status. To compound the problem, low-income black families have many children. The Bureau of the Census figures for 1974 show that black families with incomes under $10,000 had three times as many children under 18 as black families with incomes over $10,000. For whites, on the other hand, families with incomes over $10,000 had 1.3 times as many children under 18 as families with incomes under $10,000. Thus given the higher propensity of children of high-income families in general

TABLE 3—PERSONS AGED 18 TO 24 BY COLLEGE ENROLLMENT STATUS AND FAMILY INCOME, 1974

Family Income	College Enrolled[a]		Not Enrolled in College[a]		Total[b]	
	Black	White	Black	White	Black	White
Under $5,000	17	17	83	83	668	855
$5,000–9,999	30	27	70	73	525	1702
$10,000–14,999	26	37	74	63	270	2422
Over $15,000	42	50	58	50	193	4338

Source: Bureau of the Census, 1975.
[a]Shown in percent.
[b]Shown in thousands.

to enroll in college, the problem of reducing the overall disparity in the distribution of income assumes an important demographic dimension. If we accept the argument by Gary Becker that the demand for better quality (and therefore fewer) children increases with education and income, then federal incentives to greater human-capital investment in the children of families with incomes under $10,000 represent, in principle at least, a way out of the vicious circle.

Over the past decade, federal educational policy has generally been directed toward reducing the cost of formal as well as post-formal education through a wide range of tax and direct expenditure incentives. A rough measure of these incentives is provided by Congressional Budget Office data on federal aid to students. In fiscal 1977, tax expenditures and direct outlays totaled $8.5 billion. Students whose families had an adjusted gross income of less than $10,000 received 45.2 percent; 36.4 percent went to families within the income range of $10,000-20,000; 11.2 percent to families over $20,000; and the rest, 7.2 percent, to self-supporting students (see Congressional Budget Office, p. 68).

The degree to which the vicious circle may be broken by direct or indirect government subsidies must also depend on what James Meade calls "the need to accumulate." People with little capital have a greater need to accumulate capital—human or otherwise—so they will make a greater effort to do so. Indeed, immigrant groups have been able to finance heavy investment in human capital by drastically curtailing consumption. In analyzing the factors influencing the incomes of white male immigrants to the United States, Barry Chiswick has found that human-capital investments made by immigrants "depress earnings initially and raise them later on," and that "larger worker-financed investments mean a steeper post-immigration experience-earnings profile, a sharper rise of earnings with years in the United States" (p. 918). This earnings profile may very well be replicated by blacks who are induced by federal subsidies to invest heavily in human capital. That is to say, to the extent that this investment requires temporary withdrawal from the labor force, income would be reduced thus causing a temporary widening of the income gap between some groups of blacks and whites to be followed by a closing of the gap when the human-capital investment begins to bear fruit. Department of Labor projections, for example, estimate that for the period 1975–85, the civilian labor force participation rates for black males aged 20–24 years will decline from 84.2 to 82.2 percent. For black females, the projected decline is smaller, from 59.0 to 58.8 percent (see U.S. Department of Labor and U.S. Department of Health, Education and Welfare, p. 335).

How rapidly incomes rise subsequently to close the gap will be determined not only by the state of the economy, but also by the workings of the labor market. If the labor market is better explained by Lester Thurow's job competition model, then we would expect heavy investment in human capital to improve the position of minorities in the job queues of America. The closer the economy is to full employment, the faster this improvement is likely to occur. In the job competition model, the individual's place in a particular job queue is based on the relative cost to his potential employer of training him, and this in turn is determined by what Thurow calls "background characteristics," an important feature of which is education. While these characteristics determine the worker's place in the queue, it is the specific training received on the job which determines his productivity. Those higher up in the queue will get higher productivity jobs which pay higher incomes; and higher incomes provide a greater ability to acquire property which further enhances income.

If productivity resides in the job as the job competition model asserts, then as long as there are low-productivity jobs, the people who occupy them will be at the lower end of the income scale. As Michael Piore has observed, industrial capitalism will seek to fill its low-productivity jobs with workers willing to accept them wherever they can be found. Most of these jobs are created by a rapidly expanding service sector and are a

prime source of employment opportunities for the younger members of the labor force. The rate at which minority workers move out of these low-productivity jobs into high-productivity jobs will depend on the amount and quality of their human-capital investment and on the extent to which economic policy can stimulate the growth of high-productivity jobs. Thus, in the final analysis, the direction of the income disparity between blacks and whites will be determined by the complementarity of incentives to human capital investment and incentives to physical capital investment.

REFERENCES

G. S. Becker, "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries*, Universities–Nat. Bur. Econ. Res. conference series, Princeton 1960.

B. R. Chiswick, "The Effect of Americanization on the Earnings of Foreign-born Men," *J. Polit. Econ.*, Oct. 1978, *86*, 897–921.

J. E. Meade, "Determinants of Inequality in a Property-owning Democracy," in Edward C. Budd, ed., *Inequality and Poverty*, New York 1967.

M. T. Piore, "The 'New Immigration' and the Presumptions of Social Policy," *Proc., Ind. Rel. Res. Assn.*, Dec. 1974.

Lester C. Thurow, *Generating Inequality: Mechanisms of Distribution*, New York 1975.

Bureau of the Census, *The Social and Economic Status of the Black Population in the United States: An Historical View 1790–1978*, Washington 1979.

______, *The Social and Economic Status of the Black Population in the United States 1974*, Washington 1975.

Congressional Budget Office, *Federal Aid to Post-secondary Students: Tax Allowances and Alternative Subsidies*, Washington 1978.

Council of Economic Advisors, *Economic Report of the President 1979*, Washington 1979.

U.S. Department of Labor and U.S. Department of Health, Education and Welfare, *Employment and Training Report of the President*, Washington 1976.

The Persistence of Racial Inequality in Urban Areas and Industries, 1950–70

By MICHAEL REICH*

Much of the public and many professional economists believe that blacks made substantial economic gains relative to whites in the United States in the 1950's and 1960's. Some economists have even claimed that racial discrimination had already disappeared from much of contemporary economic life in the United States by the early 1970's. Recent discussion thus has not focused on whether a major movement in the direction of racial equality has occurred, but rather on whether competitive market pressures or government actions have caused the improvement.

Relative income advances have occurred, especially for black women, and significant occupational changes have taken place for black men and women. I argue here, however, that many of the gains for black males do not result from a decline in discrimination or from a decline in racial economic inequality within the urban-industrial sector of the economy, but result from the movement of blacks *between* economic sectors. The decline of a black agrarian class of small-holding and tenant farmers and the incorporation of blacks into the urban class structure has occurred in a manner that has not eliminated racial inequality, but has reproduced it in another setting.

Some economists, represented most prominently by Richard Freeman, James Smith and Finis Welch, claim that racial discrimination in the labor market has all but disappeared. Older blacks still suffer the effects of past acts of discrimination that lower their productivity and earnings, but young blacks are not so afflicted. Consequently, as young blacks, who comprise recent and new labor market entrants, replace older blacks who leave the labor market, racial inequality in earnings will also diminish. Continuing inequality is the last vestige of a now-dead pattern of discrimination. This claim has been given widespread publicity.

To support their claim, Freeman and Smith-Welch point to the decline in racial differences in returns to schooling that took place during the 1960's for young men and women. They suggest also that black-white earnings ratios have improved for young people in every schooling and family background category.

In fact, during the 1950–70 intercensal period: 1) Racial income differences narrowed *within* every age cohort, and by 1970 the historic pattern of widening racial income differences as a specific cohort grew older was no longer apparent. 2) The proportion of blacks employed in professional, managerial, and clerical occupations increased dramatically, while the proportion employed in agriculture and domestic service decreased just as dramatically. These shifts occurred at a much more rapid rate among blacks than among whites. 3) Nonwhite-white annual median income ratios rose between 1950 and 1970: from .54 to .64 for families, from .54 to .60 for males, and from .40 to .92 for females.

However, both the contention that racial discrimination in jobs has disappeared and the implication that overall racial equality may be on the horizon are not fully supported by the evidence. These statistics require qualification, for they exaggerate both the recent improvements in the relative position of blacks and the prospects for continued gains in coming years. First, much of the evidence for a decline in job discrimination is based on developments during the expansionary years of the 1960's and the

*Department of economics, University of California-Berkeley. The findings reported here are presented in greater detail, including full documentation of sources and methodology, in my forthcoming book, ch. 2.

early 1970's. More recent data reflecting the patterns of the recessionary years of the 1970's may point in a different direction. Second, Smith and Welch's own data indicate the greatest gains in the 1960's occurred for college-educated blacks. The relative gains for blacks in the lower-schooling categories, where most blacks are still situated, ranged from much more modest to nonexistent. (This pattern, together with occupational evidence cited earlier, suggests that a black professional and managerial stratum expanded significantly in recent years, but that the great majority of blacks were left behind.) Third, the growth of black professionals and managers has occurred primarily among the lowest-paying categories in these highly heterogeneous occupational groups and was concentrated heavily in public employment. Thus, in 1970 health technicians, nurses, precollege teachers, social workers, and vocational counselors accounted for 63 percent of all blacks in professional and technical occupations. In 1970 only 1.7 percent of private sector salaried managers and administrators were black; two-thirds of black managers and administrators earned less than $10,000 in 1970, compared to one-third of white managers. Fourth, since much job discrimination occurs through differential promotion and training, not through initial hiring, it is hazardous to generalize future racial differences from a few years in the labor market experiences of young workers. This latter form of discrimination may have grown in importance as overt discrimination has become illegal. Fifth, differentials in unemployment and male labor force participation rates have widened since the 1950's, suggesting that a substantial segment of the black population is falling further behind both whites and more privileged blacks. Sixth, the startling gains for black women relative to white women should be evaluated in the context of the continuing decline since 1940 in the earnings of all females relative to males as well as continuing racial differences among women in yearly hours worked and years in the paid labor force. Finally, a decline in present labor market discrimination does not imply an end to racial inequality. Many inequality-reproducing mechanisms continue to operate outside the labor market, notably class background and housing and school segregation. Freeman and Smith-Welch discuss these issues in some of their writings, but some of the inequality-generating variables remain substantially unexplored. Taken together, all these considerations cast considerable doubt on the thesis that racial discrimination is declining substantially for most blacks and that real racial equality will soon appear in the United States.

But the most telling and serious evidence against the view that racial equality is on the horizon comes from sectoral disaggregation of the national data that seem to show improved racial income ratios from 1950 to 1970. Data organized by regions, metropolitan areas and industries indicate rather the persistence of racial economic inequality for males.

A variety of regression analyses of postwar national annual data on the relative income of nonwhites indicate that while cyclical variables account for much of the year-to-year fluctuations, a positive time trend also appears significant. However, relative trends for the entire United States obscure differing trends in the regions. In every region except the South the relative income of nonwhite males shows substantial cyclical fluctuations, correlated with the national business cycle, but exhibits no upward time trend. Only the South exhibits a secular reduction in racial inequality for males. This improvement appears to begin about 1959 and appears to end by 1972. The secular national improvement noted above thus appears to result largely from both an upward trend within the South (due perhaps to migration or a possible decline in discrimination within that region) and the migration of blacks out of the South to regions that continue to have higher absolute and relative income for nonwhite males.

A similar pattern of differing regional trends appears for females. In each region except the South, nonwhite females had achieved income equality with white females by the late 1950's, though little secular

change has occurred since that time. In the South, by contrast, nonwhite female incomes rose from about .45 of white female incomes in the 1950's to about .75 in the mid-1970's. The national improvement for nonwhite women since 1960 thus seems due entirely to changes within the South and to the changing regional composition of the nonwhite population.

These regional patterns, when coupled with well-known rural-urban migration patterns, suggest the hypothesis that much of the recent relative gains of blacks result from their movement out of the rural South and into the urban and industrial areas of the North and the South.

Trends in racial inequality within urban areas in the United States support this hypothesis. Among the urban *U.S.* population, the ratio of black male income to white male income rose from .600 in 1948 to .640 in 1969. This modest urban rate of increase, while significant, was only three-fifths that for the United States as a whole.

The aggregate national urban change, moreover, obscures two opposing trends. While relative black male incomes rose in the urban areas of the South, from .533 to .605, they *declined* in the urban areas of the rest of the nation, from .726 in 1949 to .688 in 1969. By 1969, the black-white male income ratios in these two parts of the country had moved closer together, but the point of convergence was considerably *lower* than the 1949 urban non-South level. To illustrate, I have calculated future trends, assuming that the rate of change in the black-white income ratio in each region continues at the 1949–69 pace. If this assumption were to hold, the South/non-South differential would be eliminated in 1984, but at that time the urban black-white income ratio would stand at .659. This pattern does not provide a basis for optimism.

A similar picture of stagnant racial income ratios emerges in the data on individual major metropolitan areas. On the average, virtually no change occurred in the relative income of nonwhite males between 1949 and 1969 in seventeen standard metropolitan statistical areas (*SMSA*) of the United States with the largest black populations in 1970. This absence of significant improvement is found for *SMSA*s in every region of the country: The average nonwhite-white income ratio fell from .71 to .69 in *SMSA*s in the North and West, and stayed at .56 in the South.

Individual cities showed significant changes over this period, though few systematic patterns are discernible among these cities. One that may perhaps be significant is that the largest relative declines for blacks occurred in highly industrialized cities: Detroit; Pittsburgh; Birmingham.

If this twenty-year period provides any guide, it suggests that racial inequality is not diminishing within the nation's metropolitan areas. These findings lend support to the expectation that improvements in the aggregate national relative position of nonwhite males resulting from rural to urban migration may diminish substantially as the remaining rural population is depleted.

How have blacks fared in this period within individual detailed industries? Since the 1940's, black workers have entered industrial employment in unprecedented numbers. Total black employment in manufacturing has risen from 470,000 in 1940 to 998,000 in 1950, 1,306,000 in 1960, and 1,600,000 in 1970. These gains have lifted the percentage of manufacturing black employment from 0.5 in 1940 to 9.2 in 1970.

While racial employment patterns by industry have been studied, there have been few quantitative studies of trends in racial income inequalities by industry. Using decennial Census data, I have calculated nonwhite-white ratios of median annual earnings for males in the experienced civilian labor force for detailed industries from 1949 to 1969 (for the forty-nine industries in manufacturing, transportation, communications, and utilities for which the Census published relevant data in both years).

Over the period 1949–69, black employment in these industries grew from 7.4 percent to 10.4 percent of the labor force. Despite the considerable growth of black employment, the average black-white male earnings ratio in these industries increased from .71 in 1949 to only .74 in 1969. Blacks

approached equality with whites by 1970 in only two industries, taxicab service (an unusual industry in that workers have greater control over their time input) and street railways and bus lines (an industry dominated by public employment). (Preliminary investigations of the same industries in 1976, drawn from the 1977 *Current Population Survey*, indicates that these inequalities had not diminished by the mid-1970's.)

While some industries show large gains for black men over these two decades, most show very small increases (the ratio rose by .1 or more in only nine industries) and a surprising number (thirteen) show declines. Declines occurred in many major industries (such as motor vehicles and equipment or meat products) that employ a large and growing number of black workers and enjoy a relatively progressive image in race relations. Black employment as a percentage of the motor vehicle and parts labor force increased from 8.3 percent in 1949 to 13.8 percent in 1969, but the black-white earnings ratio fell from .882 to .855. Multiple regression calculations indicate that labor supply variables, such as racial differences in schooling or age, do not explain the absence of greater improvement in these 49 industries.

Relatively small increases occurred in most industries that did register an upward trend in relative earnings. In textile mill products, black employment grew from 4.9 to 12.8 percent of the total industry labor force. Yet the black-white earnings ratio remained stationary: .728 in 1949 and .736 in 1969. These examples and findings indicate that black breakthroughs into employment in the basic industrial sectors do not necessarily lead to the achievement of racial equality within those industries.

In sum, regional, metropolitan, and industrial disaggregations of national data suggest continuity rather than change in racial economic inequality in the period 1950–70. This persistence of inequality is particularly remarkable when viewed in the context of rising relative levels of black schooling attainment, cultural and political changes and changes in government policies relating to racial discrimination in this period. The stability of the disaggregated racial inequality measures in the context of the equalizing pressures of these nonmarket forces reinforces the suggestion that competition in the market has not effectively reduced racial inequality.

In addition, and also not discussed here, the growth of government employment and transfer payments has exerted upward pressure on the relative incomes of nonwhites. In 1970, the ratio of black to white male median annual earnings among government workers was .765, compared to .653 among nonagricultural private wage and salary workers. The comparable figures for females were .928 and .738. Blacks constituted a higher proportion of total government employment than of private employment and average earnings levels were higher in government employment. The growth of government employment in the 1950–70 period, therefore, worked to nonwhites' advantage. This advantage, however, has diminished considerably since 1970.

The findings summarized here suggest that insofar as the redistribution of the black labor force from the agricultural sector to the urban sector has been substantially completed, further advances in racial equality can be achieved only by challenges to the processes and structures that reproduce racial inequality within the urban-industrial setting. Standard economic theory has not yet produced a satisfactory explanation of racial discrimination and inequality. The persistence of these phenomena suggest that alternative analyses are still needed.

REFERENCES

R. Freeman, "The Changing Labor Market for Black Americans, 1948–72," *Brookings Papers*, Washington 1973, *1*, 67–132.

Michael Reich, *Racial Inequality: A Political Economic Analysis*, Princeton, forthcoming.

J. Smith and F. Welch, "Black-White Male Wage Ratios: 1960–70," *Amer. Econ. Rev.*, June 1977, *67*, 323–38.

U.S. Bureau of Labor Statistics, *Current Population Survey*, Washington 1977.

Guaranteed Employment, Work Incentives, and Welfare Reform: Insight From the Work Equity Project

By STEPHANIE WILSON, DANNY STEINBERG, AND JANE C. KULIK*

Interest in national welfare reform has generated several recent income-maintenance and employment and training demonstration projects targeted at welfare recipients and low-income persons. Income-maintenance programs typically operate by providing cash grants to poor families in amounts decreasing with earned income. Employment and training programs focus on increasing employability through skill development and job placement. While income maintenance strategies address problems of poverty and income inequality, they do not consider the capability of the disadvantaged to achieve economic independence; also, marginal tax rates on benefits have been found to decrease labor supplied (see Michael Keeley et al.). Programs employing a social services and training approach address the issues of employment and labor supply more directly. Analysis of one such program, however, suggests only slight increments to annual earnings have accrued in the postprogram period, and these only for a select group of participants (see Bradley Schiller).

Several explanations for these outcomes can be posited, one being that barriers to employability of welfare and low-income persons have not been adequately assessed or remedied. Many of these persons are unmarried female heads of household with little prior work experience, job skills, or labor market familiarity. Real barriers to employability exist for these women, some of which can be corrected via a regimen of training and improvement of existing skills, and others which are more difficult to ascertain and correct. These barriers must be overcome before individuals can be expected to benefit from any program designed to reduce their dependence on welfare. Once these barriers are overcome incentives may be used to increase their commitment to the workforce and to foster economic self-sufficiency.

The Minnesota Work Equity Project was funded as a two-year demonstration, employment, and training program serving clients from a variety of public assistance programs (AFDC, GA, Food Stamp) via a common service delivery system. This experiment addresses issues of barriers to employability and incentive to work by guaranteeing a job to all clients deemed employable according to statutory criteria. "Employable" clients are in effect required to work as one aspect of program services, though emphasis is on employability development through counseling or training.

In this paper we identify barriers to employability faced by Work Equity Program clients using preliminary data from the first months of program operation. We analyze clients' five-year preprogram work histories, determinants of the decision to work, and returns to labor market investments. We also discuss clients' reservation wage expectations which provide information about labor market orientation. Finally, we discuss policy implications of these results.

*Abt Associates Inc., Harvard University and Abt Associates Inc., and University of Chicago and Abt Associates Inc., respectively. This paper is based on a larger report prepared under contract with the U.S. Department of Labor. We would like to thank Ernst W. Stromsdorfer, principal investigator of the Work Equity Project evaluation, for advice and encouragement throughout our work on this project, and for many helpful remarks on earlier drafts of this paper. The opinions expressed here are our sole responsibility and should not be attributed to the U.S. Department of Labor or any agency or other individuals associated with the Work Equity Project.

I. Characteristics of Work Equity Program Clients

Data from a client survey (see Abt Associates) providing information on the first 975 of 2,700 persons currently enrolled in the program, shows that the typical client is an unmarried mother of young children with little previous labor market experience. Seventy percent of the sample is female, with 80 percent of these woman unmarried heads of household with dependent children. In contrast, 68 percent of male clients are married. Only 8.2 percent of clients sampled are black, and 8 percent are other nonwhites.

Women have little full-time job experience, having worked on average only 1.8 of the five years prior to program enrollment. Men have more experience and regular attachment to the workforce; on average they accrued 2.6 years of full-time experience and 37 percent worked in at least four of the previous five years.

Chronic welfare dependence is most evident among female clients. Twenty-three percent received welfare in each of the past five years, as opposed to only 1 percent of the men. Welfare payments constitute a greater fraction of total household income for female-headed families than for male; on average, benefits account for 44 percent of their total household income. Among men, the comparable proportion is 23 percent.

II. Client Labor Market and Welfare Histories

An individual's past history of attachment to the workforce and welfare system provides an estimate of future labor force behavior, in the absence of other measures. Information on length of welfare attachment, and amount and type of work experience acquired by clients during the five years prior to program referral, was gathered in the client survey; analysis of this data reveals important sex differences in patterns of participation. We focus on three characteristics which most distinguish men and women: commitment to the workforce, continuity of welfare attachment, and the interaction between work and welfare recipiency.

Men appear more strongly attached to the labor market, accumulating more years of full-time work experience and exhibiting more consistent participation in full-time employment. Of those who worked, male clients averaged 2.7 years of full-time experience. A large fraction—30 percent—was employed in all five years. Those women working averaged 1.8 years of job experience, but were much less likely to work regularly over the five-year period. A sizeable proportion of women dropped out of the labor force at some time. Women do not opt for part-time employment in proportionately greater numbers than men; 45 percent of both sexes never held a part-time job. Thus, women seem to work full time if they choose to work at all.

Chronic welfare recipiency is evident only among women. Half the women sampled have been on welfare continuously for three or more years, and almost one-fourth have received benefits in all five years. Few men have such an attachment.

Another instructive indicator is the way work and welfare have been combined in the past by clients, as this yields insight into the degree to which working may foster welfare independence. Cross classification of work experience by welfare attachment shows the two activities are often complementary among women. Approximately 18 percent of female clients who worked extensively in the previous five years received public assistance concurrently, suggesting that sample women may form two distinct groups: those working little and chronically dependent on welfare, and those working regularly yet receiving benefits which supplement earnings.

As men are much less likely to receive benefits regardless of their employment status, this analysis is less illuminating for them. Those men who work and receive assistance appear to have special problems in the labor market. Given that they possess a level of orientation that female clients do

TABLE 1—PROBABILITY OF WORKING IN THE YEAR PRIOR TO PROGRAM REFERRAL: MEAN IMPACT OF EXPLANATORY VARIABLES, MAXIMUM LIKELIHOOD PROBIT ESTIMATES[a]
(Asymptotic Normal Statistics for Probit Coefficients in Parentheses)[b]

Sample	Mean Dependent Variable	Constant	Years of School	Full-Time Work Experience Years	Age	Race Black	Married	Program Volunteer	All Children Preschool	All Children School Age	Vocational Education
687 Women	.69	.60	.01	.11	−.01	−.25	−.07	.13	−.11	.12	.08
		(.59)	(.37)	(9.51)	(−.04)	(−10.0)	(−1.90)	(15.94)	(−1.82)	(2.55)	(2.04)
288 Men	.90	.99	−.01	.05	−.05	.02	−.04	−.05	.01	.01	.08
		(2.78)	(−.26)	(11.84)	(−.42)	(4.20)	(−.76)	(−1.61)	(.32)	(.78)	(3.55)

Source: Calculated from *Baseline Survey Questionnaire*.

[a]Mean impact is calculated as the mean derivative of the estimated probability of working with respect to the appropriate explanatory variable. For the jth person and ith variable this derivative is $\hat{B}_i f(X_j \cdot \hat{B})$, where $f(\ )$ is the standard normal density, X_j is the vector of independent variables, $\hat{B}$ is the vector of estimated probit coefficients, and $\hat{B}_i$ is the ith coefficient in $\hat{B}$.

[b]Other variables in the equation: Never Married, Race-Other, Children of both school age and preschool age, children all 18 years or older. All variables except schooling, age, and work experience are dummy variables. Work Experience is years of full-time work experience in the four years preceding the preprogram study year. The omitted dummy variables are: Race-White, Divorced, No Children.

not, their behavior patterns are more difficult to explain.

III. The Decision to Work

Information describing clients' previous twelve-month labor market histories allows an examination of the importance of the preceding years of work experience and outcomes on their decision to participate and their recently earned wages. Determinants of the probability of working were estimated by maximum likelihood probit;[1] mean derivatives of this probability with respect to selected independent variables are presented in Table 1.

By far the most important determinant of recent participation is previous work experience, each year of full-time experience increasing the probability of working by 11 percentage points on average. Marital status and age structure of children were also important for women. Married women were 7 percent less likely to have worked than divorced, separated, or never married women. The child-age effects are partly due to program rules. Women with preschool children are exempt from participating in the Work Equity Project until all children reach age 6 or older.[2] Yet a number of exempt clients volunteered for the program; those volunteers having both preschool and school-age children were 7 percentage points more likely to have worked in the past year. Among mandatory participants, women with all preschool children were 11 percent less likely to have worked than women without children, while those with all children of school age were 12 percent more likely to have worked than childless women. The difference between all preschool and all school-age children is thus a substantial 23 percentage points. Vocational education also has an independent impact on probability of having worked, for both men and women, as does race.

IV. Determinants of Wage Rates and Quoted Reservation Wages

The wage earned on clients' most recently held job reflects their primary positive incentive to work and determines their prospects for earning an income above the poverty level when working full time. Currently earned wages may peg the general pay rates that program participants consider acceptable. The median wages reported during the period under study[3] were well

[1]Alternative specifications including nonearned income measures yield similar results, both in the probit equation and in the corrected regressions reported below.

[2]As all volunteers have all preschool children or a mixture of preschool and school-age children, the net effect of program volunteer status is estimated as the sum of the volunteer and the appropriate child structure dummy coefficients. For women the latter sum is .07 for mixed-child-age structure.

[3]The study period is the year preceding clients' programing referral, spanning approximately July 1, 1977 to February 1979, with beginning and end dates client specific.

TABLE 2—DETERMINANTS OF WAGES AND QUOTED RESERVATION WAGES: *OLS* REGRESSION COEFFICIENTS
(t-Statistics in Parentheses)

	Dependent Variable					
	Men			Women		
Independent Variable	*ln* Wage	*ln* Wage	*ln* Reservation Wage	*ln* Wage	*ln* Wage	*ln* Reservation Wage
Experience	.034	−.018	.001	.080	.063	.040
	(1.46)	(−.55)	(.08)	(5.30)	(2.12)	(5.17)
Age	.003	.010	−.000	.001	.002	−.002
	(0.99)	(2.30)	(−.16)	(0.35)	(0.60)	(−1.89)
Education	.015	.023	.018	.028	.025	0.04
	(1.09)	(1.65)	(2.19)	(2.38)	(1.95)	(8.24)
Vocational Education	.042	−.034	.100	.094	.078	.037
	(.753)	(−.52)	(2.88)	(2.13)	(1.52)	(1.75)
Race-Black	.158	.140	.049	.122	.178	.027
	(1.35)	(1.19)	(0.67)	(1.49)	(1.48)	(0.82)
Married	.140	.176	.045	−.031	−.018	−.087
	(1.72)	(2.14)	(0.86)	(−0.49)	(−.26)	(−2.79)
Never Married	−.036	.017	−.150	−.083	−.080	−.070
	(−.33)	(0.15)	(2.29)	(−1.51)	(−1.45)	(−2.64)
Selection Bias		−.742			−.142	
		(−2.28)			(−0.64)	
Number of Children Aged < 18			.06			.031
			(4.36)			(4.12)
Intercept	.977	.930	1.18	.630	.733	.668
	(4.72)	(4.50)	(9.36)	(3.86)	(3.19)	(8.96)
R^2	.097	.116	.27	.118	.119	.194
S.E.E.	.177	.174	.077	.176	.177	.057
N	251	251	288	452	452	679

Source: Regression analysis of data discussed in text.

above the 1979 federal minimum wage, at $4.00 per hour for men and $3.21 for women. However, 26 percent of the women and 14 percent of the men reported earning less than $2.65 on their last job.

Various estimates of the contribution of experience, schooling, and training to wages are presented in Table 2. It might be expected that members of a welfare population would reap smaller than average returns to investments in the labor market, as they may be more likely to have taken jobs offering little training or opportunity for advancement.[4] Our data strongly rejects such hypotheses for women. In simple wage equations, returns to education are well within the range of recently reported estimates for the overall population, at 2.9 percent per year of schooling. Returns to work experience are higher; at 7.9 percent per year of full-time experience. Vocational training is estimated to yield over 9 percent higher wages. Following James Heckman and G. S. Maddala, sample selection correction by entering a Mills ratio based on probit coefficients from Table 1 lowers these estimates somewhat but does not alter general conclusions.[5] Our best prediction for the entire sample of women, including those not working in the past year, is that experience contributes 6 percent to wages, and vocational education, 7.8 percent.

Estimated returns for men are smaller and in a simple wage equation only marital

[4]Alternatively we expect low estimated returns as the sample is selected on unfavorable labor market outcomes.

[5]This specification and others and the peculiarity of the negative estimated coefficient for the bias term are discussed in greater detail in Steinberg and Kulik.

status is significant at conventional levels. These estimates may be effected by sample selection, here, restriction to a welfare population. However, if we accept these estimates, then neither increased work experience nor training can be expected to improve the wages of welfare men.

All clients sampled, whether they worked in the past year or not, were asked the lowest wage at which they would accept employment.[6] Table 2 reports estimates of a simple reservation wage equation, including both standard wage offer and reservation wage determinants. Responses for both sexes are similar to, but more uniform than, those of actual wages. The equation standard error is half of that in the observed wage equation. Results from the wage offer and reservation wage equation strongly suggest minimum wage voluntary public works programs such as EOPP may face recruitment difficulties.

V. Implications

Several tentative inferences can be made on the basis of the above observations. Sample clients can be divided into distinct subgroups having unique employability difficulties. These subgroups include women with little previous work experience and continuous welfare attachment, women with substantial work experience and some welfare dependence, and men with no apparent barriers to employability but who experience difficulties in the labor market. Our analysis suggests chief barriers to employability for the first group consist of lack of familiarity with the labor market, absence of marketable skills, and low earnings potential. Given our estimates, work experience has the largest reliable effect on wages; providing this group with fully subsidized jobs would simultaneously provide exposure to the working world and experience which could be translated into higher earnings.

Slightly different considerations pertain to the second group of women. While program sanctions may ensure that these women work more while enrolled, they will not necessarily encourage increased work effort after program completion. Career counseling and training coupled with the subsidized job may aid in these clients' obtaining a job with earnings potential such that they could reduce their dependence on public assistance still further.

At this point we remain agnostic regarding implications for the men. The men are less likely to have observable barriers to employability, yet our analysis indicates that wage-rate responses to increased investments are negligible.

Programs requiring nonexempt welfare recipients to work, provided they also render support services, may contribute significantly to the degree of economic independence achieved by welfare women. If the deterrents to labor force participation for these women have indeed been lack of skills, work familiarity, and sufficient incentive, then those inexperienced women with children constituting the largest subgroup of welfare recipients should benefit most. If the program yields benefits for men, these will probably come in the form of increased hours worked and job tenure, rather than on the wage front. As the program has been operational slightly over a year, outcomes will not be measurable with confidence for an additional twenty-four months.

[6]While self-reported "reservation wages" are not supported by behavior, they provide evidence regarding the accuracy of client-wage expectations.

REFERENCES

J. Heckman, "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals Econ. Soc. Measure.*, Fall 1976, *5*, 475–92.

M. Keeley et al., "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs," *J. Hum. Resources*, Winter 1978, *13*, 3–36.

G. Maddala, "Self-Selectivity Problems in Econometric Models," unpublished manuscript, Univ. Florida, Sept. 1976.

B. R. Schiller, "Lessons from WIN: A Manpower Evaluation," *J. Hum. Resources*, Fall 1978, *13*, 502–23.

Danny Steinberg et al., *The Work Equity Project Study Sample: Client Profile and Labor Market History*, Cambridge, Mass., Dec. 1979.

Abt Associates Inc., *The Minnesota Work Equity Research Project Baseline Survey Questionnaire*, Cambridge, Mass., June 1978.

Rates of Return to Human Capital: A Test using El Salvador Data

By LASCELLES ANDERSON*

Most of the empirical work on human capital has been confined to conditions in developed countries. This is probably due to difficulties in obtaining data for other countries. It would be interesting to see how robust the theory and its associated models are in conditions regarded as typical for developing countries. This will be considered here using a large data set describing the El Salvador labor market.

Two questions will be examined, namely, (a) How well does the earnings function of the human capital theory explain earnings distribution in El Salvador; and (b) Does the return to training, called experience here, approximate the return to school as is often assumed in human capital research? The latter issue is of course related to the notion of "overtaking," the point in the experience profile when the flat earnings stream is intersected from below by the earnings stream affected by postschool investments.

I. The Data

The basic data for this exercise is derived from a large data file of merged statistics on individuals in the El Salvador labor force. In all, 33,000 observations are included in the data base. The file was generated for use in a major study of education planning being carried out at Harvard University Graduate School of Education through the Center for the Study of Education and Development. From that file, I selected out those cases, 5,999 in all, which would allow for the usual test of the importance of human capital variables on earnings distribution. The sample consists of males who were at least twenty years of age in 1975. The variables selected were age, income, and years of education. Mean years of education are just over seven years, and using Mincer's definition of experience, mean years of experience are twenty-six years approximately. More importantly, however, and consistent with the implications of the human capital model, mean earnings rise with years of education. In the regression analysis which follows, I assume that those with one year of education are for all practical purposes illiterate. The results indicate that mean earnings for illiterates are a little more than half that for persons with fifteen or more years of education. These are, however, not starting salaries and so can be assumed to include an unknown experience component.

It is interesting that these mean earning figures rise monotonically over the education range. Regression results, later identified, suggest that these profiles are nonintersecting for the most part.

II. Human Capital Relationships

The basic human capital model is concerned with life cycle phenomena: investments in human capital phased over one's entire life, and the pattern of earnings which is related to such a path of investment expenditures. The choice is to invest so as to maximize the returns to that investment, subject to the condition that the time during which benefits can be derived gets shorter as the investment period is extended. Yoram Ben-Porath pioneered this kind of analysis (see also Gary Becker). The earnings function is a good summary of the way in which the small set of human capital variables —schooling, experience and sometimes, weeks worked—affect the distribution of earnings.

Assuming that labor markets function perfectly, that there are no inequalities in

*Associate professor of education, Harvard Graduate School of Education.

intelligence, physical skills, or home background, and that all persons have equal access to capital markets, as well as making certain assumptions about training and training costs, most importantly the assumption that education involves postponement of entry into the labor force, the empirical earnings function can be written simply as

$$(1) \qquad \log Y_s = \log Y_0 + rS + u$$

where Y_s is earnings at time s, S is years of education, and u is a disturbance term. Equation (1) assumes that other variables, collected in the disturbance term, for example age, are not systematically related to $\log Y$. This is not likely to be the case. But while it might appear that age should enter as an explanatory variable, it is experience that human capital theory identifies as crucial. The model estimated in this paper has several variants. Consistent with (1) above, I estimated four separate regression equations:

$$\text{(a)}\ \log Y_s = f_1(S)$$

$$\text{(b)}\ \log Y_s = f_2(S,T)$$

$$\text{(c)}\ \log Y_s = f_3(S,T,S^2)$$

$$\text{(d)}\ \log Y_s = f_4(S,T,S^2,(S\cdot T))$$

where T is years of experience, S^2 is the square of S, and $(S\cdot T)=(S)\times(T)$.

The results clearly indicate that schooling explains approximately 30 percent of the dispersion in relative earnings, and the rate of return is just over 17 percent. When experience is included among the independent variables, the explanatory power of the model rises slightly to approximately 31 percent, and the return to education also rises to just under 18 percent. Both variables are significant. Equation (c) shows no real change in the explanatory power of the model by the inclusion of the square of years of education, a variable that would identify the effect of increased education at the margin, and the variable does not enter significantly. When an interaction term is added in equation (d), however, the rate of return to schooling falls to 15 percent, but the explanatory power of the model rises to 32 percent. Accordingly, although the effect of the interaction term is small, the variable enters significantly.

Briefly, then, schooling does seem to matter, and the rate of return for the sample as a whole ranges from 15 to 19 percent. Experience also appears to matter; the effect of education seems to decline at the margin, and there appear to be significant interaction effects between education and experience, though these effects are relatively weak.

III. Education Effect at Overtaking

One important result of human capital theory associated with the work of Jacob Mincer is the argument that due to post-school investment decisions of individuals, there will be a tendency for dispersion in earnings of persons with a given amount of education but varying amounts of post-school investment in human capital, to increase at first, and decline, to cross at a certain point and diverge again. The point is crucial since persons with the same level of education are assumed to choose occupations which lead to similarity in lifetime incomes. Due, then, to varying profiles of post-school investment, the convergence of earnings profiles takes place at the point at which those profiles representing differing post-school investments cross. Mincer (1974) calls this point the "overtaking" point, and estimates that it occurs approximately seven to nine years subsequent to entry into the labor force. Since this overtaking point is then, by the logic of the model, the point at which the explanatory power of schooling is greatest, it is of some interest in any test of the human capital earnings function model.

In order to derive the point of overtaking, say $\hat{j}$ years, Mincer used a variety of heroic assumptions in order to generate a flat earnings profile, Y_s. Such a flat profile excludes post-school investment magnitudes and is thus a pure schooling effect profile. Since

post-school investment initially depresses the earning profile of investors, Y_0 where the respective profiles cross, as at $\hat{j}$ years of work experience, can be identified with the point of overtaking. But the flat earnings profile used here has never been observed, and it has to be assumed that the rate of return to schooling is the same as the rate of return to post-school investment to derive that profile. Note, however, that some circularity of reasoning seems to creep in here. Because of that, the effect of schooling at "overtaking" is derived in this paper by standardizing for experience, following Mincer (1975) and G. Psacharopoulos.

To derive these results, regressions were run with the *log* of income as the dependent variable and schooling as the independent variable, for different five-year-experience categories. I derived the results for the twelve experience categories. The regression equation for each experience category is

$$(2) \qquad \log Y_s = \log Y_0 + rS + u$$

The results should have been broadly consistent with the assertion that the amount of relative dispersion explained by schooling alone becomes greatest at the overtaking set. However, although a first peak in rates of return to schooling appears at the five to nine-year experience category, explained variance is not greatest there. Additionally, derived rates of return are inconsistent with the theory in its asserted relation between the overtaking period and the rate of return to schooling. Rates of return vary from a low of 8 percent for those at the extremes of the experience spectrum, to 35 percent for those within the forty-five to forty-nine-year experience category. The R^2 rises to nearly 15 percent between twenty and twenty-nine years of experience and falls after that, with the one exception of the secondary peak of 13.8 percent at the forty-five to forty-nine-year experience category. A search for possible explanations for such deviation of results from the theory turns up the following. If the explanatory power of schooling should be greatest at the overtaking point, not only should the R^2 rise, but the explained variance should also be large. The results here show explained variance of less than 15 percent, when much higher levels should have emerged. One explanation for this is that the data was reported, not by years of schooling for each observation, but by schooling range, as for example, three to five years. There was no way of getting actual schooling levels, the variable for each observation being taken by the midpoint of the range given for that case.

Such data limitations severely restrict the variability in the schooling variable, and thus most likely accounts for the substantially reduced power of the education variable in the finer analysis done within experience groupings. The overtaking results here cannot therefore be regarded as an adequate test of the theory, given data limitations.

IV. Effect of Experience Without Schooling

A large number of the cases in the sample are cases with little or no education. It is of some interest to see what the rate of return to training or experience is likely to be under such circumstances. To accomplish this, a portion of the sample was thus selected out, and regressions run, relating $\log Y$ to experience, in effect holding schooling constant at zero years of education. The resulting rate of return thus is a pure experience rate, to be compared to a pure schooling rate. The equation estimated was

$$(3) \qquad \log Y = \log Y_0 + rT + u$$

The result was

$$(4) \qquad \log Y = 2.2638 - .00631\,T \atop \qquad\qquad\qquad (0.0013)$$

The surprising thing that emerged was that experience entered with a negative sign, but the variable was significant. One plausible interpretation of this result is that there is a bias in the El Salvador labor market for younger workers, among those who are illiterate. This could be explained by employer feelings that possibility for training is greater when workers are younger. Recall that the mean level of education is fairly low. Under such circumstances, training is

likely to be training on the job, and it is a fair expectation that younger workers would be given the edge, both in hiring as well as in pay increases, based on some measure of training or retraining potential. Another likely explanation is that the nature of the work could generate such results. If the jobs are those that require physical strength and illiterates have these jobs, younger workers might be more productive.

The result, which appears unusual in human capital terms, could have an alternative explanation in terms of the economics of information. If education functions more as a screening device used by employers, then in labor markets where education levels are uniformly low, some other means for the identification of productive potential must be used. Our results could suggest that for experienced illiterate workers, managers have sufficient information concerning productive potential, but for younger workers, such information is not yet available. An inverse relation between experience and *log* earnings could thus suggest the working of such a mechanism in labor markets like that of El Salvador.

The negative sign on the experience variable for illiterate workers indicates the possiblity of a dual structure in the El Salvador labor market and suggests that the assumption sometimes made in human capital theory at least in its empirical work, of equality between training and schooling rates of return, may not be a very reasonable assumption in all labor markets.

V. Conclusion

The results of this analysis suggest that returns to schooling in El Salvador lie approximately in the 15 to 19 percent range, and that these rates rise with education levels. The analysis suggests, however, that schooling rates may not approximate the same rates of return to training in labor markets where mean levels of education are low, and where illiterate or otherwise inadequately schooled workers are a large part of any random sample of the economically active population, as is the case in the El Salvador sample studied.

REFERENCES

Gary S. Becker, *Human Capital*, New York 1975.

Y. Ben-Porath, "The Production of Human Capital and the Life-Cycle of Earnings," *J. Polit. Econ.*, Aug. 1967, *75*, 352–65.

Jacob Mincer, *Schooling, Experience and Earnings*, New York 1974.

———, "Progress in Human Capital Analysis of the Distribution of Earnings," in Anthony B. Atkinson, ed., *The Personal Distribution of Income*, Boulder 1975.

G. Psacharopoulos, "Schooling, Experience and Earnings," *J. Develop. Econ.*, Mar. 1977, *4*, 39–48.

Theodore W. Schultz, *Investment in Human Capital*, New York 1971.

How Interested Groups have Responded to a Proposal for Economic Competition in Health Services

By ALAIN C. ENTHOVEN*

In the spring of 1977, President Carter reaffirmed his support for universal and comprehensive national health insurance. That summer, as a consultant to HEW Secretary Joseph Califano, I designed a proposal based on fair economic competition of alternative health care financing and delivery plans. The proposal is called Consumer-Choice Health Plan (*CCHP*).

Briefly, today's exclusion of employer health benefit contributions from employee incomes subject to income and payroll taxes would be replaced by a refundable tax credit usable only as a premium contribution toward a qualified health care financing and delivery plan. A qualified health plan would be one obeying certain rules outlined below. In the full version of *CCHP*, the tax credit would equal 60 percent of the average per capita cost for covered services for people in each demographic category. The present Medicare system would be replaced by a system of premium subsidies in fixed dollar amounts, equal in real value to the average cost to Medicare today of people in each demographic category. Medicaid would be replaced by a system of premium subsidies in addition to the tax credits. The subsidies would be on a sliding scale, inversely related to income, and equal to 100 percent of actuarial cost in the case of the very poor. Competing private health plans would quote premiums by demographic category. People in higher medical risk categories would pay higher premiums (so that health plans would be adequately compensated for serving them) and receive correspondingly higher subsidies.

To be a recipient of these tax credit and other subsidy revenues, a qualified health plan would have to comply with the following rules. First, it would have to participate in an annual *open enrollment* in which it would have to accept all enrollees who chose it, without regard to age, sex, race, religion, national origin, or prior medical conditions. The government would manage the process in a manner analogous to that used in the Federal Employees Health Benefits Program (*FEHBP*) in such a way as to assure that everybody was given a full and free choice. Second, each plan would practice *community rating*, charging the same premium to all persons in the same demographic category enrolled for the same benefits. Third, all plans would have to insure a common set of basic health services. They would have to offer one option limited to such services. They would also have to provide "catastrophic expense protection," limiting each family's annual out-of-pocket expenses for basic health services. Finally, they would have to participate in a program of information disclosure designed to facilitate informed consumer choice.

Consumer-Choice Health Plan was not adopted by the Carter Administration. While the plan attracted considerable support from the economists in the administration, the leaders of the Department of Health, Education, and Welfare were opposed to it. The latter were already fully committed to a regulatory approach to cost control, as exemplified by the administration's twice-defeated Hospital Cost Containment proposal.

*Graduate School of Business, Stanford University.
©Institute for Contempory Studies.

One response to *CCHP* was Congressional and Domestic Policy Council inquiries concerning low-cost incremental proposals designed to foster economic competition in health services and the growth of alternative delivery systems without instituting a full universal health insurance scheme. In February 1979, I presented the following proposals to Congressman Al Ullman, Chairman of the Ways and Means Committee, and subsequently to other congressional and Executive Branch leaders. First, as a condition for continued favorable tax treatment of employer-provided health benefits, every employer above a certain size would be required to offer its employees a choice of no fewer than three competing health care financing and delivery plans meeting certain standards. Second, the employer's contribution would be the same whichever plan the employee chooses. (Today, most employers do not offer their employees an economically fair choice of health plan. Either they offer no choice, or they offer a choice but pay more if the employee chooses a more costly health plan.) Third, all qualifying plans would include coverage of standard basic minimum benefits, "catastrophic expense protection," and continuity of coverage provisions. I also recommended a proposal previously developed by leaders of the Health Maintenance Organization (*HMO*) industry and subsequently sponsored by the Carter Administration. This proposal would permit each Medicare beneficiary to direct that the "Adjusted Average Per Capita Cost" to the Medicare program for people in his actuarial category who are not members of an *HMO* be paid in the form of a fixed prospective payment as a premium contribution to the *HMO* of his choice. This would require Medicare to contribute equally on behalf of people choosing fee-for-service and *HMO*s. (Actually, the Carter Administrations' proposal was for a premium contribution to *HMO*s equal to 95 percent of equality, to permit the government to share in the savings.) Congressman Ullman, Senator Durenberger of the Finance Committee, and other key congressional leaders have endorsed the above proposals and introduced legislation to put them into effect. The rest of this paper is a summary of the response to these proposals by each of the main interested groups.

What follows are personal observations and not the result of a systematic survey. Moreover, none of the groups is monolithic; each includes people with various views. And the positions of members of each group change with changing perceptions of what is the most clear and present threat to their interests. Moreover, some groups profess to support competition in principle, but oppose attempts to implement it. So interpretation is needed. I attempt to describe and explain what I understand to be the predominant view of each group in 1979.

I. Organized Labor

The labor movement has a long-standing commitment to universal social insurance, including a single comprehensive standard of health care for everyone as a matter of right. In the labor view, services should be provided on the basis of medical need, and neither the cost of care nor ability to pay should be a factor in the patient's use of services. Union leaders do not share the economists' view of the market as an efficient or appropriate allocator of resources. They instead share the traditional "liberal" (in the American sense of the term) faith in government as the instrument of desirable social change.

While the principles of Consumer-Choice Health Plan are not incompatible with universal health insurance at a high—if not single—standard of comprehensive care, *CCHP*'s emphasis on diversity and choice appears to labor leaders to be in conflict with an emphasis on raising a single standard of comprehensive health insurance. At best, if *CCHP* is not incompatible with universal health insurance at an acceptable standard of care, it is seen as a diversion from the main struggle.

Moreover, the rhetoric of competition and incentives for efficiency is unpalatable to the labor point of view. The point that the purpose of the proposals is to make

doctors compete to serve *workers* does not appear as a major redeeming factor. The idea that it is possible to take some aspects of resource allocation out of the market (for example, by not allowing insurers to practice preferred-risk selection), while leaving other aspects in the market, is a fairly subtle and recent idea, and not likely to be very appealing to someone who mistrusts markets to begin with.

In collective bargaining, the traditional goal of labor has been to have the employer pay for 100 percent of comprehensive health benefits. Here, social philosophy has been reinforced by the tax laws which make such benefits tax free to the employee. On the way to this goal, health benefits have provided labor leaders with an apparently inexhaustible source of bargaining prizes with names that have emotional appeal. Thus, it has been very important for labor leaders to be able to bargain for and control entitlements to employer-paid health services. Also, solidarity in the bargaining process depends on everyone being treated alike. Lots of individual options can diffuse the force of a bargaining position. These ideas conflict with the notion that each employer should offer a fixed-dollar contribution toward the health plan of the worker's choice and that each worker should be offered a multiple choice of health plans.

Both *CCHP* and the bills introduced by Congressman Ullman and Senator Durenberger include changes in the tax laws that would put limits on the amount of tax-free employer contributions to health insurance. But some unions are among the groups that receive the largest employer contributions. Such groups would be almost bound to lose from a change in the tax laws that gave an equal tax subsidy to all families.

Labor leaders do like prepaid group practice but dislike the principle of fee-for-service. This confronts them with a dilemma. Traditionally, they have fought for comprehensive benefits, as opposed to limited benefits with high coinsurance and deductibles. This has been favorable to prepaid group practices which traditionally offer comprehensive benefits. It is difficult for prepaid group practices to compete against limited insurance plans with high deductibles and coinsurance. Where consumers have such a choice, those expecting few medical expenses will take the low-premium insurance, leaving the prepaid group to suffer from adverse risk selection. In recent years, however, the benefits covered by negotiated insured fee-for-service plans have become quite comprehensive, and the premiums often exceed the premiums for prepaid group practice plans. The practice of requiring the employer to pay 100 percent of the premium has come to mean forcing employers to subsidize fee-for-service against prepaid group practice.

Labor leaders will have to find a way to modify their position in order to resolve the dilemma. One indication of such a change can be found in the new Kennedy plan, the Health Care For All Americans Act, which labor supports. In that proposal, consumers would be offered a limited choice of health plans, and health plans that can deliver comprehensive benefits for a community-rated premium less than the government-mandated ceiling would be allowed to make cash rebates to their enrollees as a way of attracting their business.

II. Business

There has not been a comparable unified view among business leaders. Businessmen understand the principles of fair economic competition, and generally approve of them as long as they are not applied to their own industry. Some employers have taken a positive view toward health plan competition. For example, in the early 1970's, a group of employers in Minneapolis–St. Paul studied the problem of health care costs and decided to encourage competition. In 1970, there was one Health Maintenance Organization in the area with 36,000 members. By 1979, there were seven with a combined membership in excess of 300,000. The support and encouragement of employers was a decisive factor in that development. Leading employers in that area offer their employees a multiple choice of health plans. Some employers elsewhere have also embraced the concept of health plan competition. Some

have adopted the policy of offering their employees the opportunity to join *any* federally qualified Health Maintenance Organization serving their area. However, for the most part, employers have not applied the principles of fair economic competition to their own purchases of health care services.

Several factors appear to contribute to the dominant employer practice of not offering choices. First, most still think of traditional insured fee-for-service as the normal and appropriate way to buy health care services. They think of alternative delivery systems such as prepaid group practice, individual practice associations, and primary care networks as something suspect. Indeed, the idea of health plan choices only makes sense in the context of alternative delivery systems which have not existed in most areas until recently. Some employers would readily offer alternative delivery systems if they could provide care at substantially less cost. But the alternative systems are not yet large enough to realize economies of scale.

Employee benefits managers are often concerned that offering choices means more administrative work for them, or upsets existing insurance arrangements. Most employers accept the markets for health services and insurance as they are, and seek their own individual best interest, rather than seeking to bring about fundamental structural reform that would be better for all employers in the long run. There is an element of Prisoner's Dilemma in the situation. Employers use health benefits as a tool for attracting and retaining employees. Thus, they want to be able to say they offer a "better" health plan, not merely an opportunity to buy membership in the same health plans offered by other employers.

Many employers focus on the premium cost only, and not on the total of premium and out-of-pocket costs. Health Maintenance Organizations often have lower total costs but higher premium costs because their premiums cover comprehensive services.

There is some benefit to the individual employer from offering choices. The employees can get more consumer's surplus from the employer's dollar if they are given distinctive choices. But the main benefits of competition come with the lower rate of growth of health care costs when most employees in an area are offered a fair choice. Although health care costs may now be a large item, a company must perceive that it can gain a competitive advantage before it will be motivated to invest resources in health care cost reduction. Although economic competition is likely to benefit all employers in a community, the incentive to be the first to lead the way is not strong.

Some employers regret having agreed to pay for 100 percent of comprehensive service benefits on an open-ended fee-for-service basis. But having done so, they would find it exceedingly difficult to take away benefits to which their employees feel entitled. But if they offer a choice, and 100 percent employer payment of either premium, the employee is likely to have little or no economic incentive to choose the less costly plan. Some employers are beginning to resolve this problem through the use of "Cafeteria Style Benefit Plans" described in the Revenue Act of 1978, under which employees choosing the less-costly health plan are allowed to apply the savings to other benefits. However, if the employer is already offering a choice and 100 percent payment of either premium, then such equalization of employer contributions means either "take away" from one group of employees or an additional outlay in the short run.

At the level of national legislation, it is still too soon to tell how business will respond to procompetitive legislation. Some business leaders have already endorsed an approach to health care cost control based on incentives and competition in the private sector. The National Chamber of Commerce has supported this competitive approach. Other business organizations and leaders are likely to do so. However, the force of their support will be attenuated by other legislative priorities. On the other hand, some business leaders see any requirement to offer employees choices as an unwarranted intrusion into management prerogatives.

III. Medical Profession

The medical profession has traditionally opposed economic competition in health care services. For example, it has insisted on the principle that every insurance plan should provide for "free choice of doctor." This principle is the opposite of the principle that every consumer should have the right to limit his choice of doctor to those participating in a closed panel plan, voluntarily, in exchange for what he perceives to be better benefits or lower costs. The medical profession has resisted attempts by insurance plans to discriminate between participating providers who accept cost controls and nonparticipating providers who do not, and it has backed its resistance with boycotts of "limited provider" insurance plans.

Today, the medical profession faces a dilemma. The noncompetitive insured fee-for-service system is clearly associated with the rapidly increasing cost of health care services, which is causing a serious financial problem for government. The main response of government has been in the form of public utility-type controls on prices and utilization. Most doctors dislike the prospect of such detailed bureaucratic controls even more than the prospect of economic competition.

In 1976, the Board of Trustees of the American Medical Association established an independent National Commission on the Cost of Medical Care. In 1977, the Commission recommended economic incentives in purchasing insurance and health plans (offering employees choices, employer contributions that are equal with respect to the health plan chosen, equal tax treatment), consumer cost sharing, fair market health plan competition, and alternative financing arrangements. In June 1978 the House of Delegates of the American Medical Association gave a limited endorsement of these recommendations. Other physicians also have favored economic competition. In September 1978, the American Association of Foundations for Medical Care—the association of physicians involved in Individual Practice Associations—endorsed Consumer-Choice Health Plan.

There appears to be a great deal of ambivalence about competition in the leadership of the AMA and among physicians in general. Of course, the short-term economic interest of the medical profession is in maintaining the status quo, which may also be the AMA's goal. Their 1979 lobbying efforts were focused on defeating the Carter Administration's proposed Hospital Cost Containment bill, and not on formulating and backing legislative proposals to enhance economic competition.

IV. Hospitals

The status quo also has important advantages for hospitals. They now face a growing demand that is not price sensitive. The widespread use of cost reimbursement relieves their managements of the burden of hard economic choices. But this situation also has important disadvantages. Government cost-containment efforts have focussed on hospital spending as the largest and one of the fastest growing components of total health care spending, resulting in an increasingly complex web of regulation. Costs under Medicare are reimbursed but subject to a complicated set of limits and disallowances, so that reimbursements fall short of actual costs. A constant struggle by both hospitals and government to bend the accounting rules to their respective benefit has produced a situation of great complexity. New investments cannot be made without Certificate-of-Need, and this process has proved to be very expensive, time consuming, and highly political.

In the past, some hospital industry leaders have favored state rate regulation in the belief that they could be the dominant influence on the regulatory agencies. They saw regulatory agencies as setting rates that would cover their costs and protect them from attempts by some third-party payors to pay less than full cost. However, industry leaders saw the Carter Administration's Hospital Cost Containment proposal as likely to subject hospitals to permanent de-

tailed and often perverse federal controls. Hospitals find themselves caught between the growing demand fueled by insured fee-for-service and the government's cost control efforts.

Leaders of the hospital industry know that prepaid group practices and some other types of health maintenance organizations hospitalize their patients 25 to 45 percent less than similar people cared for under insured fee-for-service. So economic competition among such organized systems is likely to produce a considerable reduction in the demand for hospital services. However, industry leaders also know that hospitals will always be needed, that some of the demand for inpatient services would be replaced by demand for hospital outpatient services, and that hospitals would be able to adapt to competition in many ways. Generally speaking, hospital administrators look forward to a leading role in a more organized health care delivery system.

While none of the national hospital associations has yet taken a formal position on the Ullman and Durenberger bills, the general position of the leadership of these associations has been quite positive towards competition. While this might be viewed as a stratagem to help defeat the Carter Administration's Hospital Cost Containment proposal, I believe that the support for a strategy based on economic competition runs deeper.

It is apparent that hospital industry leaders recognize that the government will be forced to do something very fundamental about controlling costs. They generally take the long view and recognize that their choice is between detailed federal controls and a system of true economic competition. In the latter situation, hospital administrators would be facing local private sector buyers such as health maintenance organizations, conscious of quality as well as cost, informed about local conditions, and empowered to make agreements based on judgment—all quite the opposite from what they would face in federal price control authorities. Under economic competition, they would be rewarded for finding ways to provide better care at less cost. I have found considerable support for that idea among hospital administrators.

V. Commercial Health Insurance Industry

The commercial health insurance industry is opposed to proposals to create economic competition in health services such as *CCHP* and the incremental proposals described above. Its spokesmen generally phrase their opposition in code words such as "impractical" or "an administrative nightmare." Further, industry spokesmen say they would want to be sure that the choices offered are *real* choices. I take this to mean that it is acceptable to require employers to offer Health Maintenance Organizations because *HMO*s are few in number, usually small, and under present conditions are not likely to become a serious threat to commercial insurance. But it is not acceptable to make insurers compete with each other in the same employee groups.

The industry view is that there is too much competition now. Therefore, they seek an antitrust exemption to allow them to band together to negotiate with providers over fees and charges. In fact, there is a great deal of competition in health insurance. Insurers compete with each other and with employer self-funded insurance for the exclusive business of an employer. But all plans insure for essentially the same services from the same providers. They compete on their administrative costs but are virtually powerless to control which services are rendered and hence total costs. Competition among insurers therefore does not create competition among providers, and only the latter offers hope of bringing about changes toward less costly styles of care.

The industry favors public utility regulation of hospitals, in the form of Certificate-of-Need controls on capital investment and federal or state controls on hospital rates and budgets. Not only do they endorse existing control systems, but they also recommend legislation to extend them and to make the controls more pervasive. One reason the health insurance industry favors

state rate regulation is because they see in it an opportunity to secure equality among payers. Today, Medicare and Medicaid often reimburse less than hospitals' full costs. In some states, Blue Cross' strong market position and historical relationship with hospitals enable Blue Cross to obtain a substantial discount. The result is that the hospitals make up the difference by higher charges to private paying patients, the beneficiaries of the insurance companies, putting the insurance companies at a competitive disadvantage.

There are exceptions to the industry position. Some insurance companies are innovating and positioning themselves to be able to succeed in a competitive world. One of the most notable examples is SAFECO Life Insurance Company of Seattle which has developed a primary care network *HMO* in which participating physicians are paid a per capita retainer fee for providing all office-based primary care services to enrolled patients, and a cost-control incentive payment for managing all their specialist and hospital care. Prudential has started group practice *HMO*s, and has even teamed up in Dallas with Kaiser-Permanente to create a new *HMO* there. Insurance Company of North America, which owns hospitals and *HMO*s, has testified forcefully in favor of economic competition.

VI. Blue Cross-Blue Shield

The Blue Cross-Blue Shield Association (*BC-BS*) is a trade association representing 143 independent not-for-profit *BC-BS* plans. There is great variety in the history and market positions of these organizations, so one cannot easily generalize about them. One can not expect plans with the very high market shares some enjoy to favor proposals requiring employers to offer choices. However, Walter J. McNerney, President of the *BC-BS* Association, has consistently testified in favor of the Federal Employees Health Benefits Program as a model for national health insurance. In that program, employees are offered a multiple choice of private health plan and a fixed-dollar contribution by their employer.

One might wonder what differentiates *BC-BS* from the commercial insurance carriers. Both do most of their health insurance business on the basis of insured fee-for-service, a financing system that would seem likely to lose market share under a system of fair economic competition. However, the insurance carriers have traditionally been quite remote from the providers of care, and have paid for health services through indemnity payments and major medical insurance. Blue Cross and Blue Shield, on the other hand, were sponsored, respectively, by hospital and medical associations. And although they have gradually become more independent of provider control, they have remained much more deeply involved with providers. While competition would cause insured fee-for-service to lose market share, *BC-BS* could more easily develop alternative delivery systems than could commercial insurers. In fact, they now are participating in about 66 alternative delivery system developments. Thus, *BC-BS* leaders see themselves as likely to succeed in a system of fair economic competition.

VII. Conclusion

At the time the Federal Employees Health Benefits Program was being considered by Congress in 1959, each of the interested groups had its own ideas as to what the program should look like. What finally emerged was a compromise that all groups could accept. It happened to be a model of fair market competition. That such a program could simultaneously embody rational economic principles and a workable political compromise inspires some optimism about its potential as a model for universal health insurance.

Health Insurance and Cost-Containment Policies: The Experience Abroad

By UWE E. REINHARDT*

Private health insurance coverage in the United States has grown extensively during the past several decades. Roughly 94 percent of expenditures for hospital care are now covered by third-party payment, and 61 percent of expenditures on physician services. While coverage for some items remains thin—for example, only 17 percent of expenditures on drugs are now covered, and only 20 percent of dentists' services—private and public insurance programs now cover an overall average of about 70 percent of all outlays on personal health care (U.S. Department of Health, Education, and Welfare (*HEW*), Table 162). The equivalent figure in 1950 was only about 35 percent (U.S. Department of Commerce *Statistical Abstract...1977*, Table 105).

By American standards, these numbers are impressive. They are less so by international standards. Most developed nations now offer their citizens far more comprehensive coverage, and look upon the American health insurance system as something of an oddity. Foreigners familiar with our system, for example, are puzzled how a modern industrial nation can remain content to leave an estimated 10 to 25 million of its citizens without any health insurance coverage whatsoever.[1] They are equally amazed that a decade after the introduction of the much acclaimed Medicare program, retired persons in the United States still pay directly some one-third of the cost of their health care in the form of copayments or premiums for supplemental coverage.

In assessing the current debate over national health insurance in this country, it may be useful occasionally to view our system in a broader context. To provide such a perspective, the present essay will describe the health insurance systems in three major industrial nations: Canada, France, and West Germany. The objective will be to tease from that synopsis several questions American health economists might fruitfully explore.

I. Health Insurance in Selected Nations

Table 1 summarizes the main features of the Canadian, French, and West German health insurance systems. In each of the countries virtually the entire population has comprehensive coverage for personal health-care expenditures. Cost sharing on the part of patients is either absent or very modest.

A. Development and Structure

National health insurance in Canada is really an agglomerate of independent provincial insurance plans. In each of the ten provinces, coverage for hospital care is administered by a plan distinct from that for physician services. The hospital plans were established in 1958 under the federal Hospital Insurance Diagnostic Services Act. Comprehensive coverage for physician services was introduced during 1968–71 under the federal Medical Care Act of 1966, although the Province of Saskatchewan had such a program as early as 1962.

Under the two acts, the federal government agreed to share roughly half of any provincial insurance plan that met a set of stringent federal guidelines. According to

*Professor of Economics and Public Affairs, Princeton University. Research for this paper has been supported by Ford Foundation Grant No. 730-0094, and is gratefully acknowledged.

[1]According to the Social Security Administration, some 24 to 25 million Americans did not have any health insurance coverage in 1976. The Congressional Budget Office estimates that the number of uninsured is only 11 to 18 million. In a study by Pracon, Inc., the number was put at 12 million. In this connection, see Health Insurance Institute, p. 20.

TABLE 1—OVERVIEW OF NATIONAL HEALTH INSURANCE SYSTEMS

Parameter	Canada	France	West Germany
Administration	Ten independent provincial medical and hospital plans operating within federal guidelines and with federal cost sharing.	A national system of sickness funds organized on a geographic basis and supervised by the Ministry of Health and Social Security.	A mosaic of over 1,500 sickness funds organized in state and national associations and operating within federal statutes.
Role of Private Insurance	Confined to provision of supplemental coverage.	Confined to provision of supplemental coverage.	Covers some 7 percent of the population for basic benefits.
Population Coverage	Public system provides universal coverage on equal terms.	The national system now covers 99 percent of the population.	The statutory system covers 93 percent of the population. Some 7 percent has private coverage; 1 percent has no coverage.
Benefits	Medical services, hospital services, for special groups: dental care, drugs.	Medical services, hospital services, prescription drugs, medical appliances, cash benefits.	Medical services, hospital services, dental care, prescription drugs, medical appliances, cash benefits.
Predominant Mode of Reimbursing Providers	Hospitals: global budgets Physician: fee-for-service reimbursement under province-wide, negotiated fee schedules.	Hospitals: per diems plus fee-for-service for physician services Hospital physician: salary Private physician: fee-for-service.	Hospitals: per diems Hospital physician: salary Private physician: fee-for-service
Financing	A mixture of taxes and direct premiums. Mixture varies from province to province. Federal govt. bears about 50% of costs.	Basically a payroll tax; about 3.5% of income paid by employee; about 12.5% paid by employer.	A payroll tax of about 11% of earnings shared equally by employer and employee.
Cost Sharing by Patients	Originally rare. Increasing somewhat in provinces that permit physicians to opt out of the system.	Coinsurance rate of between 30% to zero, depending on severity of illness. Average coinsurance rate about 10%.	There is virtually no cost sharing except for a modest copayment per prescription.
Percent Growth Health Expenditures average annual increase for period indicated	1971–74: 12.1 1971–75: 18.0 1975–77: 13.0[d]	1970–75: 16.8[c] 1975–78: 15.9 1978–79: 16.8[d]	1970–75: 20.1[a] 1975–76: 9.8[a] 1976–77: 4.8 1977–78: 7.1[d]
Percentage of Gross National Product Going to Health Care	1975: 7.1 1976: 7.2 1977: 7.1[d]	1975: 6.8[c] 1978: 7.1 1979: 7.3[d]	1976: 9 to 10[b]

Sources: Available from the author upon request.
[a]Outlays under the statutory health insurance system only.
[b]Rough estimate. A figure comparable to the *U.S.* "national health expenditures."
[c]The French figures are akin to the *U.S.* "personal health care" series.
[d]Preliminary estimates.

the Medical Care Act, for example, the provincial plans must be publicly administered, be universally accessible on equal terms to all residents, and remove all financial barriers to covered services. Substantial cost sharing on the part of patients is inconsistent with the spirit of federal legislation.

Because responsibility for health care falls within the domain of the provinces, this peculiar federal-provincial arrangement is the closest politically feasible approximation to a bona fide *national* health insurance scheme in Canada. The cost-sharing mechanism is also the main vehicle through which the federal government can exert control over aggregate health expenditures. In 1977, for example, the federal government began to limit the growth of its contribution to the growth in gross national product, leaving the provinces to absorb a greater share where aggregate outlays grew at a faster rate. More recently the federal government has decided to make its contribution not solely in cash, but substantially in the form of tax points that vacate a share of the federal tax base in favor of the provinces. More direct forms of cost control are thus fully in the hands of the provinces, as will become apparent further on (see Table 3).

The West German health insurance system is a network of about 1,500 fiscally and administratively independent "sickness funds" (*Krankenkassen*) whose common element is, once again, that they operate within a framework of strict federal statutes. The

sickness funds trace their origin to the friendly (cooperative) societies established during the late nineteenth century by unions or other associations of workers. They were forged into a national system in 1883 when membership in a sickness fund was made compulsory for certain groups of blue collar workers.

The main thrust of the early sickness funds was to provide their members with cash benefits during episodes of illness and with only a limited range of medical services. Since then, emphasis has shifted from cash benefits towards comprehensive medical benefits in kind, and an ever larger proportion of the population has come under compulsory membership, or has been granted the right to join the statutory system.

The premiums for insurance coverage under the statutory system are based on the insured's ability to pay rather than on actuarial principles. As of August 1979, the premium covering an employed person (along with dependent family members) ranged between 7 to 13.8 percent of the employee's gross income, with half that premium being paid by the employee and the other half by his or her employer (*Die Ortskrankenkasse*, p. 821). Coverage for retired persons is paid by the latters' pension funds although at rates below actuarial costs. At this time only about 7 percent of the population (mainly civil servants entitled to generous federal cash indemnities and the self-employed) are covered by truly private commercial insurance (see Ulrich Geissler).

Although West Germany's sickness funds are controlled by strict federal statutes they are nevertheless private, not-for-profit organizations that are expected to assure their own fiscal stability. It may be thought that the very number of them—about 1,500—must introduce stiff competition for members. Actually that number is deceptive. West Germans typically have only limited freedom in the choice of their sickness fund, because membership tends to be dictated by the insured's occupation and/or geographic location. Furthermore, the federal statutes impose upon all funds a common benefit package that is so extraordinarily broad as to leave only modest room for competition through benefits.

Some occupation groups, notably white-collarworkers, do have the option of joining a so-called *Ersatzkasse* (Substitute Fund) instead of the fund their employers would otherwise dictate. Persons who join the statutory system voluntarily also may elect one of the *Ersatzkassen* which form an integral part of the statutory system. The *Ersatzkassen* may offer voluntary members cash indemnities rather than service benefits—a right denied the other funds. The *Ersatzkassen* also have tended to pay physicians higher fees and are therefore thought, by the insured, to provide superior physician services. Finally, the *Ersatzkassen* were instrumental in forcing the conversion of the erstwhile capitation reimbursement of physicians (predominant until the late 1960's) to the current fee-for-service method. Such limited competition as there has been among West German sickness funds has therefore tended to favor the economic position of physicians. Since the *Ersatzkassen* are chosen voluntarily and have registered a growing membership, their impact seems to have been favored by at least some proportion of the insured. Depending upon a statutory system's prevailing position, then, competition among insurers may increase or decrease the cost of health services.

Like the West German system and many other systems on the European continent, the French system is composed of a network of sickness funds that trace their origin to cooperative societies of industrial workers. Over time, however, the system has come to be dominated by the so-called "General Scheme" (*Régime Générale*), a regionally organized but centrally directed network of some 120 local, primary sickness funds (*caisses primaires d'assurance maladie*) that covers most of the country's industrial work force and their dependents (thus, altogether, some three-quarters of the French population). The primary sickness funds of the *Régime Générale* are supervised by 16 regional funds (*caisses regionales d'assurance maladie*), which in turn are coordinated and

TABLE 2—AVERAGE ANNUAL GROWTH IN NATIONAL HEALTH EXPENDITURES: CANADA AND UNITED STATES, 1970–75

		Canada		United States	
National Health Expenditures,	1975	Can$	11,372	US$	122,231
in Billions of Current Dollars	1970		6,081		69,201
Average Annual Growth in					
National Health Expenditures[a]			13.3		12.1
National Health Expenditures					
Per Capita, in Constant	1975	Can$	350	US$	407
(1970) Dollars	1970		286		334
Average Annual Growth in					
Constant-Dollar Health					
Expenditures Per Capita[a]			4.1		4.0
Average Annual Growth in the					
Price Index of:[a]					
Medical Care (United States)					
Health and Personal Care (Canada)			6.3		6.9

Sources: Available from the author upon request.
[a]Shown in percent.

supervised at the national level by the National Sickness Fund Association (*Caisse National d'Assurance Maladie*). The latter is a public corporation whose task it is to assure the fiscal viability of the entire network of funds through the setting of premiums, the definition of benefit packages, and the extension of grants and loans from a central fund to individual local sickness funds. At the apex of this system stands the Ministry of Health of the central government which now has considerable statutory authority over the system, including the approval of negotiated fee schedules. The thrust of recent legislation has been to fold the remaining distinct groupings of sickness funds (for example, funds covering agricultural or railroad workers) into the *Régime Générale* and thus to convert the French system into a truly uniform, centrally directed national health insurance system.

B. Control over Costs and Expenditures

Table 1 indicates that the rapid secular growth in aggregate health care expenditures during the 1970's is not a uniquely American phenomenon. The experience has been shared by most other industrialized nations—certainly by Canada, France, and West Germany. Indeed, an enduring conclusion from the numerous recent international conferences on the "health-care cost explosion" has been that the phenomenon has so far been rather insensitive to the particular financial scheme grafted onto the health-care delivery system. This point is brought out even more clearly in Table 2, where American data are contrasted with comparable Canadian data, after adjustments for differentials in inflation rates and population growth. It is striking how similar the growth in real per capita expenditures has been, in spite of the great differences in these two nations' health insurance systems.

During the first part of the 1970's, the growth of health-care expenditures everywhere tended to be merely observed with some concern. Concern gave way to alarm sometime during the mid-1970's. At about that time, most nations began attempts to curb the growth in expenditures through overt public intervention. The form of this intervention has varied from country to country, in line with the institutional framework through which such intervention must work. The explicit or implicit goal of these policies, however, has been identical: to peg the growth of national health expenditures to the growth of gross national product, at least over the long run. That goal is, of course, rather arbitrary, for there is no em-

pirical support for the proposition that a society either should or would prefer to allocate a constant proportion of its resource budget to a commodity such as health care.

Policies to reduce aggregate expenditure on health care have been aimed at various combinations of the following targets: (a) *Total health-care expenditures*, 1) Total expenditures on particular sectors (for example, ambulatory care), 2) Total revenues of particular providers (for example, hospitals); (b) *Utilization of health services*; (c) *Prices of health services*, 1) Supply and use rate of inputs (manpower, beds, etc.), 2) Prices of inputs, 3) Organization of the health-care production process; (d) *Premiums for health insurance*. The targets could be attained indirectly through the provision of suitable financial incentives. A preferred approach seems to have been more direct regulation.

A hypothesis with wide currency among American economists is that an appropriate degree of expenditure-containment could be achieved simply by assuring competition in the market for health services and by confronting the actors in these markets—consumers, producers, and third-party payers—with appropriate financial incentives. On this hypothesis public policy needs not be targeted directly on any particular good, but merely needs to assure a framework for the appropriate determination of targets *b* and *c*, thus obviating the need for policies such as Certificates of Needs for hospital facilities (target *c.1*). The sine qua non of such a framework is thought to be competitive markets for health services—the ostensible goal of current endeavors by the Federal Trade Commission—and substantial cost sharing on the part of patients. A more indirect approach to the development of the framework would be the establishment of truly competitive health insurance markets, an approach best exemplified in the Consumer-Choice Health Plan proposed by Alain Enthoven.

Health-care providers in the United States—and elsewhere, for that matter—typically profess support for the concept of the competitive-market approach, if only because that approach draws away fire from targets closer to home (for example, targets *c.1* and *c.3*). Unfortunately, the providers' enthusiasm for the model tends to wane once the meaning of "competition" is explained to them more concretely within the context of their own markets. Physicians, for example, generally show little enthusiasm for competition from self-employed paramedical manpower, for an expanded supply of medical manpower, or for policies to inform consumers better with regard to both the price and the quality of medical treatments.

American policymakers seem even less impressed by the potential of competitive health-services and health-insurance markets. Although this set of decision makers cannot be characterized by one view, actual health policy in this country does betray a penchant for more direct forms of intervention—be it direct constraints on physical inputs (target *c.1*) under the Health Planning and Resources Development Act (1975), constraints on health-care utilization (target *b*) through Professional Standards Review Organizations (*PSROs*), unilaterally imposed limits on fees and per diems (target *c*) under the Medicare and Medicaid programs, or the repeatedly proposed caps on hospital revenues (target *a.2*).

This skepticism toward the market model is shared by policymakers in other countries as well. Canada, France, and West Germany, for example, do not rely at all on the notion of competitive markets in their approach to cost and expenditure containment. Broadly speaking, the main thrust of their policies has been to constrain expenditures on physician services through negotiated or imposed fee schedules, and expenditures on inpatient services through controlled per diems of global budgets, on the one hand, and controls on physical capacity on the other. Table 3 presents a highly condensed summary of the main cost-control instruments used in the three countries. Of the three countries, only West Germany has actually introduced a formal cost-containment law—the federal Health

TABLE 3—DOMINANT ELEMENTS IN THE CONTROL OF HEALTH-CARE EXPENDITURES

Target	Canada	France	West Germany
	A. Expenditures on Physician Services		
1. Fees	Fairly effective provincial control through negotiated schedules.	Fairly effective control through negotiated schedules that require central government approval.	Negotiated fee schedules with compulsory arbitration, but without requirement of explicit government approval.
2. Utilization of services	Controlled to some extent through monitoring of physician profiles.	Controlled to some extent through monitoring of physician profiles.	Controlled to some extent through monitoring of physician profiles.
3. Supply of physicians	No formal policy to limit supply strictly, although immigration of MD's is discouraged.	No formal policy to limit supply for purposes of cost control.	No formal policy to limit supply for purposes of cost control.
4. Cost sharing by patients	No intention as yet to introduce it formally, although opting out by physicians introduces an element of cost sharing.	There is a modest degree of cost sharing but it is not viewed as a workhorse of cost containment.	No intention as yet to introduce it.
5. Global caps on aggregate outlays	None	None	It has been attempted in recent years to agree on a total outlay for physician services and prescription drugs *ex ante*.
	B. Expenditures on Hospital Care		
1. Revenues	The individual hospital's revenues are predetermined through budgets that must be approved by the provincial government.	Control is attempted through negotiated per diems that require the approval of the *Departement's Prefet*.	Control is attempted through negotiated per diems that require the approval of the state government which can, in principle, set the per diems.
2. Utilization	No direct control	No direct control	No direct control
3. The capacity of hospitals	Controlled through regional planning by the provincial governments which finance roughly two-thirds of capital outlays.	Controlled through regional planning supervised by the central government.	Because capital expenditures are financed by a mixture of state and federal funds, there is authority at the state level to limit capacity through regional planning.
4. Prices of hospital inputs	No direct control	No direct control	No direct control
5. Cost sharing by patients	None	Virtually none	None

Care Cost Containment Act of 1977. The overall thrust of this legislation is to constrain the growth of expenditures through annually negotiated predetermined budgets at the aggregate level.

None of the three countries assigns a significant role to patients in the cost control process. There remains a modest degree of coinsurance in France (Table 1), but its average impact has declined over time and is likely to do so in the future. Canada and West Germany have adopted first-dollar coverage as a matter of principle. Although physicians in both nations have called for cost sharing on the part of patients, neither consumers nor their legislative representatives have so far shown any inclination to move in that direction.

Physicians typically make the case for cost sharing on the arguments that it would elicit more responsible conduct on the part of patients, would free medical practice from trivial cases, and contribute toward expenditure containment. It is doubtful whether physicians seriously believe the expenditure containment argument. Organized medicine is not known to favor policies that reduce the aggregate flow of funds to physicians. As M. L. Barer et al. have argued, a more plausible explanation for the profession's posture is that cost sharing, coupled with third-party coverage, is believed by physicians to draw more fiscal nourishment to the physician sector than could otherwise be had from third-party payers under universal first-dollar coverage. They may well be right.

Just why patients and policymakers in other nations are so reluctant to embrace the notion of cost sharing is even more intriguing. Economic theory suggests that a rational consumer would be unlikely to elect first-dollar coverage at actuarially fair premiums. Although universal national health insurance tends to sever the link between the actuarial cost of insuring the individual

and the premium imposed on him or her, it is not clear why a majority of individuals would expect to benefit actuarially from tax-financed universal first-dollar coverage. Perhaps one should look for other explanations.

While Americans have gradually taken to the precept that access to *adequate* health care is a basic human right, Canadians and Europeans are more committed to the notion that all technically available health care should be available for all citizens on an equal footing. Although a progressive coinsurance scheme meeting this requirement could be devised, it may be felt that in practice the incidence of coinsurance would be unlikely to be progressive and thus violate the principle of equal access.

The typical American health economist recognizes that the upper-income classes in any nation tend to enjoy, *de facto*, superior access to health care and education, but would typically argue that this ought to be so as a matter of principle. One senses from discussions with policymakers and policy analysts in other nations that they would like the structure of their health insurance system to convey at least the impression of an egalitarian distribution of health care, while no such pretense is as yet made in the United States.

There is also the possibility that traditional economic analysis tends to misspecify the consumer's objective function and thus explain normal behavior as irrational. As Victor Fuchs has recently suggested, consumers may desire health insurance not only to reduce uncertainty, but also to obviate the need for troublesome moral choices during episodes of illness. If the sole purpose of insurance coverage were reduction of uncertainty, the optimal (utility maximizing) coinsurance rate would not generally be zero. Preference for a zero coinsurance rate is much more readily explained by what Fuchs calls a penchant for "pre-commitment" (p. 170) a conscious decision to avoid the necessity of weighing money prices when making choices on the use of health care.

A penchant for precommitment, as Fuchs points out, could also explain a preference for the simple and compulsory insurance schemes that seem to be preferred in Canada and Europe. Such schemes may be preferred not only because choosing among alternative insurance schemes—each with a different benefit package and cost—is time consuming and thus costly. There is also the probability of regret over choice of an option that has turned out, *ex post*, to have been inappropriate.

II. Concluding Remarks

In our public debates on health policy, the prospect of *universal* health insurance coverage is typically viewed as an inherently undesirable but probably unavoidable compromise of free market principles. Most proposals for national health insurance have therefore sought to minimize interference with private markets for health services and health insurance, subject only to the attainment of some stated, minimally adequate degree of health insurance coverage. From an economist's perspective, this approach seems appropriately cautious and rational. Its main drawback has been that the fusion of an ethically acceptable insurance program with free market principles turns out to be a complex task whose practical—as distinct from conceptual—accomplishment has so far eluded policymakers.

Policymakers in most other nations seem to have despaired long ago of the feasibility of preserving classical market mechanisms for socialized commodities, that is, commodities to which individuals have a right regardless of income. The thrust of policy in West Germany, for example has been not to rely on classical market mechanisms, but to develop instead new quasi markets that fall somewhere between classical, atomistic markets and centralized administration favored in France. (In this connection, see, for example, Philipp Herder-Dornreich.) The ideal decision-making units in these quasi markets are freestanding associations of the individuals and organizations active in these markets (for example, association of providers, of insurers, of the insured, and so on), and the atomistic decision making in classical markets gives way to collective

bargaining among the freestanding associations, all within a statutory framework that guards the rights of weaker parties and provides for compulsory arbitration of inconclusive negotiations.

At this time neither the "classical market" approach favored in the United States, nor the "quasi market" approach favored in West Germany, nor the more centrally directed systems of Canada and France can safely be judged a resounding success, for neither model has yet had a chance to be fully operative. Whatever approach one favors, however, one may find it instructive to keep an eye on the others.

REFERENCES

M. L. Barer, R. G. Evans and G. L. Stoddart, *Controlling Health Care Costs by Direct Charges to Patients: Snare or Delusion?*, Ontario Economic Council, Occas. Paper No. 10, Toronto 1979.

A. C. Enthoven, "Consumer-Choice Health Plan," *New England J. Med.*, Mar. 23; 30, 1978, 650–709.

V. R. Fuchs, "Economics, Health and Post-Industrial Society," *Health Soc.*, Spring 1979, *57*, 153–82.

U. Geissler, "Health Care Cost Containment in the Federal Republic of Germany," mimeo., Nov. 1978.

Philipp Herder-Dornreich, *Strukturwandel und Soziale Ordnungspolitik*, Cologne 1977.

***Die Ortskrankenkasse*,** Nov. 1, 1979, *61*.

Health Insurance Institute, *Source Book of Health Insurance Data 1978–79*, Washington 1979.

U.S. Department of Commerce, *Statistical Abstract of the United States 1977*, Washington 1977.

U.S. Department of Health, Education, and Welfare, *Health—United States 1978*; H.E.W. Publ. No. (PHS) 78–1232, Dec. 1978.

Stabilization Policy and Capital Formation

By ROBERT E. HALL*

Every recession in the *U.S.* economy calls forth proposals for remedial stimulus. The government's own expenditures on goods and services can be increased, additional income can be provided to consumers in the hope that they will spend more, new incentives can be provided for investment, and the money stock can be increased. Conventional analysis does not distinguish among these policies with respect to the ratio of their effects on output and inflation. Each can push the economy out of recession, back to full employment, at the cost of worsening inflation. All operate along the same Phillips curve.

My purpose here is to reconsider the prevailing dogma by examining the possibility of differential effects of stabilization policies operating through capital formation. Expansionary policies either favor investment (monetary expansion and investment incentives) or discourage it (increased government expenditures or consumption). The paper begins with a bit of evidence to support the view that the Keynesian expenditure multiplier is positive but less than one, so that increases in other categories of aggregate demand depress investment. It also presents somewhat more conclusive evidence that monetary expansion stimulates investment. I conclude tentatively that the conventional instruments of expansionary policy do provide us with an important choice about the participation of capital formation in the resulting expansion.

The paper then asks what difference it makes for the economy whether a pro- or anti-investment strategy is pursued in response to a recession. Does the likely difference in the flow of investment over the course of the typical recession make enough difference in the capital stock so that changes in productive capacity are a matter of concern? I find that a pro-investment policy that actually restores full employment in the face of a recession will leave the economy with as much as 7 percent more capital by the end of three years than will an anti-investment policy that also restores full employment. Full-employment output is more than 2 percent higher after a pro-investment stimulus. The key hypothesis underlying this conclusion is the high marginal product of capital inferred from the earnings of capital in the current *U.S.* economy. Some reasons to think capital earnings exceed the marginal product of capital are mentioned as possible reasons for weakening the conclusion.

The conclusions also require modification when applied to stabilization policies of the magnitudes politically realistic in the United States. Antirecessionary policy has been timid and even perverse in the postwar decades. The principal source of foregone capital formation has been our failure to do anything about recessions, not our active use of anti-investment stimulative policies.

I. Evidence about the Effects of Expenditure and Monetary Policies on Investment

Though a number of large-scale econometric models of the *U.S.* economy were constructed or are used actively for exactly the purpose of predicting the effects of alternative stabilization policies, the results of these models are now greeted with extreme skepticism among macro economists. The models contain many purportedly

*Hoover Institution and department of economics, Stanford University, and National Bureau of Economic Research, Inc. This research is part of the National Bureau of Economic Research's program in economic fluctuations and is supported by a grant to the NBER from the National Science Foundation.

structural equations whose actual function is primarily to express the autoregressive behavior of their dependent variables. The result, when the models are applied to issues of policy interventions, is to grossly overstate the sluggishness of the economy. Consequently, interest has turned to direct estimation of the reduced-form relations between the endogenous variables of the economy and the exogenous variables of policy. Robert Barro (1977, 1978) has been a leader in this line of research; he has studied unemployment, real output, and the price level. So far as I know, there have not been any studies of investment within the reduced-form approach. The results presented here are of course subject to the very basic criticism that they rest on the hypothesis of exogeneity (see, for example, Stephan Goldfeld and Alan Blinder). If monetary and expenditure policies have been motivated by something other than a desire to offset movements in the economy as they occur, then we can learn the effects of policies on investment simply by regressing investment on variables expressing the magnitudes of the policies. At the other extreme, if policies have been carefully tailored to eliminate all unwanted movements in investment, there may not be any regression relation, even though policy has profound effects on investment. Because policy has been far from perfect by any standard in the postwar period, because in any case it is clear that policy moves have been extremely timid when they were explicitly countercyclical, and because presumably it is output and employment, not investment, that is the principal target of policy, I think it is interesting to examine the reduced-form evidence for investment, even though I recognize that it is not fully convincing.

In the results below, expenditure policy is measured by total government expenditures (including state and local) in constant dollars of 1972 (G). Apart from trend, the largest movements in this variable occurred during the two military adventures of the postwar period. Expenditures are measured on the conceptual basis of the national income and product accounts, as government use of goods and services. Transfers, which are part of the budgets of governments, are not part of expenditures measured in this way. The other policy considered here, monetary policy, is measured by the aggregate amount of currency and demand deposits (M_1) deflated by the implicit deflator for GNP. The use of money in real terms is justified for this purpose by the findings of Barro (1978) and others that the influence of the money stock on the price level takes two years to be felt at all. Within that period, exogeneity of the nominal money stock and exogeneity of the real money stock amount to the same thing. Finally, other determinants of investment (in the reduced-form sense) are characterized by a constant, a linear time trend, and a stochastic residual. Obviously, the residual contains the effects of many other important economic mechanisms. The validity of the reduced-form approach does not rest on the inclusion of all relevant variables in the regression. Rather, what is important is the lack of correlation of the residual with the two policy variables, that is, their exogeneity.

As a measure of investment, I have used gross private domestic investment in constant dollars of 1972. It is important to note the inclusion of capital formation other than business fixed investment in this measure —inventories and residential housing are the most important of these. The regression results for quarterly data over the period 1954 to 1979: 2 are

$$(1)\quad \underset{(73)}{I_t = 4} + \underset{(.40)}{1.59t} + a_0G_t + \ldots + a_{15}G_{t-15} + b_0M_t + \ldots + b_{15}M_{t-15}$$

$R^2 = .9785$; $s = 5.8$ *billion* 1972\$; *rho* $= .70$

The distributed lag coefficients are shown in Table 1. According to these results, government expenditures have a distinctly negative impact on investment for the first few quarters. Though investment responds slightly positively to expenditures of a year or two earlier, the negative response to the most recent expenditures is sufficiently strong that the net response to a sustained increase in expenditures is negative. The

TABLE 1—DISTRIBUTED LAG COEFFICIENTS OF G AND M IN EQUATION (1)

Years	a_i	Standard Error	b_i	Standard Error
0	−.31	.20	.69	.29
1	−.17	.08	.75	.11
2	−.05	.09	.61	.13
3	.03	.10	.37	.14
4	.08	.09	.10	.11
5	.09	.07	−.15	.08
6	.08	.07	−.34	.08
7	.04	.08	−.43	.10
8	−.01	.08	−.44	.10
9	−.07	.07	−.35	.09
10	−.12	.06	−.20	.08
11	−.14	.08	−.02	.11
12	−.13	.09	.12	.14
13	−.06	.09	.15	.13
14	.09	.07	−.01	.09
15	.34	.17	−.48	.27
Sum	−.31	.34	.37	.46

reduced-form evidence for the effect of government expenditures on total real GNP suggests it is somewhat positive, but the multiplier is less than one. The private components of GNP, especially investment, have a negative relation to government expenditure. Unfortunately, the sampling variability in these estimates is sufficiently large to preclude any strong conclusions about the exact magnitude and timing of the effects of expenditures on investment.

For monetary policy, the results are much stronger. The response of investment to an increase in the real money stock is vigorous and immediate. An increase of \$1 billion in the money stock raises investment in the same quarter by \$690 million. If the increase is sustained into a second quarter, investment is higher by \$1.44 billion; the effect peaks at \$2.52 billion for an increase sustained for five quarters (the sum of the first five lag coefficients). Thereafter, the influence lessens as the coefficients turn negative, reflecting the accelerator mechanism at work. A permanent increase in the real money supply, even if it could be achieved, would have little effect on the level of investment in the long run. Moreover, since prices eventually respond in proportion to an increase in the nominal money stock, simple monetary expansion is incapable of raising the real money stock permanently. But the evidence does support the hypothesis of a strong temporary effect of monetary expansion on capital formation.

I tentatively conclude from this evidence that government policy does have two instruments for expanding aggregate employment and output with very different implications for investment. Stimulus of aggregate demand through increased government expenditures appears to discourage investment, while stimulus through monetary expansion encourages investment. In fact, this is probably the major path by which money influences aggregate output. In the face of a negative shock on aggregate demand from some outside source, policy can restore full employment through either expenditure or monetary policy. The choice between the two should depend on, among other things, social views about capital formation.

II. Investment, Capital Stock, and Output

Now I want to assume that a suitably vigorous use of expansionary policy (or policies in tandem) has offset what otherwise would have been a typical recession set off by an adverse shock to demand. The economy faced an episode of diminished

investment as well as diminished output; in fact, investment as a ratio of *GNP* invariably falls sharply in a recession. The policy choices I wish to examine are:

1) *Pro-Investment*. Use monetary expansion or a policy with similar effects to restore investment to its normal relation to output. Specifically, investment is to be 15 percent of *GNP*, its postwar average in nonrecessionary years.

2) *Anti-Investment*. Use increases in government expenditures to restore full employment but depress investment even below its low level of a recession and early recovery. Specifically, investment is to be 10 percent of *GNP* in the first year, 11.7 percent in the second year, 13.3 percent in the third year, and its normal 15 percent in succeeding years. By comparison, investment was 13.6 percent in 1954, 12.9 percent in 1958, 13.7 percent in 1961, 14.4 percent in 1970, and 11.9 percent in 1975. It has never been as low as 10 percent during the postwar period, but similarly, vigorous expenditure policy has never been used to head off an incipient recession.

The full-employment growth economy described by Robert Solow provides a natural framework to compare these two full-employment policies. I will make use of the following assumptions and notations: 1) The labor force grows at a constant rate n; 2) Labor-augmenting technical progress occurs at rate v; 3) Capital depreciates at rate d; 4) In year t, gross investment is a fraction s_t of gross output; 5) Production is governed by a Cobb-Douglas production function and the elasticity of output with respect to capital is b; and 6) Capital per efficiency unit of labor in year t is k_t.

Then, according to Solow's model, the capital-labor ratio evolves in the following way over time:

$$(2) \quad k_{t+1} = (s_t k_t^b + (1-d)k_t)/(1+n+v)$$

A reduction in this year's rate of investment, s_t, reduces next year's capital stock and so reduces full-employment output next year, in an amount governed by the elasticity of output with respect to capital, b. In succeeding years, capital is also lower on account of diminished investment this year even if the investment rate returns to normal next year and later. The proportional effect of a reduction in the flow of investment on the stock of capital depends essentially on the sum of the rate of depreciation d, the rate of labor force growth n, and the rate of technical progress v. The higher is the sum, the more leverage current investment has on the capital stock in percentage terms.

Of the various parameters of this model, the most critical is the elasticity of output with respect to the capital stock b. Under competitive conditions, this parameter should equal the ratio of the gross earnings of capital to the gross value of output. In 1978, this ratio was 0.30, with gross earnings of capital estimated as the sum of capital consumption allowances, corporate profits before tax, net interest, and rental income plus one-half of indirect business taxes plus one-third of proprietors' income. An elasticity of 0.30 means that full-employment output is reasonably sensitive to changes in the capital stock, sensitive enough so that transitory stabilization policy may well influence output in a noticeable way. The reader should be cautioned, however, that direct examination of the data on output, capital stock, and other factor inputs has not confirmed an elasticity as high as 0.3—estimation of aggregate Cobb-Douglas production functions with postwar *U.S.* data has yielded elasticities around zero. Students of "potential *GNP*" have tended to omit capital and base their measures solely on labor input (for example, George Perry). Further, there is an unexplained gap between the apparent earnings of capital and the cost of capital in debt markets. In recent years, earnings, appropriately corrected for measurement biases due to inflation, have been robust, while interest rates, adjusted for inflation and taxation, have been zero or negative. Either businesses have failed to pursue investment to the point of equating the marginal earnings of capital to the cost of financing the capital (essentially the hypothesis of this paper), or measured capital earnings includes components other than the competitive return to capital. If profits,

proprietors' income, and rents include a large and growing element of returns to entrepreneurial talent, intangible investments, and monopoly income, then my estimate of the elasticity of output with respect to the capital stock is overstated.

For the other parameters, I have used the following estimates: 1) The rate of growth of labor productivity v is 1 percent per year, in line with the disappointing experience of the 1970's. 2) The rate of growth of the labor force is 1.5 percent per year. 3) The rate of depreciation of capital, averaged over equipment, structures (including residential), and inventories, is 10 percent per year.

With this preparation, it is possible to compare the evolution of the economy under a pro-investment response to its evolution under an anti-investment response. I will assume that the economy starts out at its steady-state capital-labor ratio with an investment rate of 15 percent of gross output. The pro-investment response maintains the investment rate at this level, so the economy continues on its growth path at 2.5 percent per year increase in output. The anti-investment response pushes the investment-*GNP* ratio to 10 percent in the year of the shock, 11.7 percent in the next year, 13.3 percent in the following year, and then it returns to 15 percent. The consequences of the anti-investment policy can be expressed in terms of the percentages by which its output and capital stock fall below those in the pro-investment, steady-state economy, as shown in Table 2.

TABLE 2—EFFECTS OF ANTI-INVESTMENT POLICY

Year	Investment Rate	Percent Shortfall in Output	Percent Shortfall in Capital
0	10.0	0.0	0.0
1	11.7	1.2	4.1
2	13.3	2.0	6.4
3	15.0	2.2	7.2
4	15.0	2.0	6.6

With a continuation of the investment rate at 15 percent in subsequent years, the shortfalls in output and capital will gradually decline to zero, though the perceptible effects of the temporary reduction in investment will continue for at least a decade. The adverse effect of the shock and the anti-investment response peaks in the third year, when the shortfall in output is 2.2 percent. The average growth rate of output from year 0 to year 3 in the economy with the anti-investment response is 1.9 percent per year, as against 2.5 percent per year with a pro-investment response. Over the period, the anti-investment policy deprives the economy of about a quarter of its potential growth over the three years.

An interesting implication of this finding is that the stimulus to capital formation from monetary expansion offsets part of the inflationary influence of the extra money. Consider, for example, the use of monetary stimulus to avoid a recession which would have depressed the investment rate to 12 percent from its normal level of 15 percent. The results presented in the previous section suggest that about a 2.5 percent increment in the money stock is needed to raise the investment *GNP* ratio by 1 percentage point. Closing the gap of 3 percentage points would require a 7.5 percent jump in the money stock. Eventually, this will make its way into higher prices, but the extra capital stock resulting from the pro-investment policy means that the long-run difference in the price level compared to the economy with no countercyclical policy will be almost 2 percentage points lower. To put it another way, toleration of a recession with its sharp reduction in capital formation is inflationary in the longer run, because the reduction in potential output calls forth a higher price level for any given money stock. Similarly, treatment of a recession with an anti-investment policy is also inflationary, again because the price level depends in the longer run on the ratio of the money stock to output.

The policy moves presupposed by this analysis are extremely aggressive by the standards of postwar experience in the United States. Raising the money stock by 7.5 percent to head off a recession is not a policy that would ever be considered seriously by the Federal Reserve. Year-to-year

variations in money growth rarely exceed 2 percentage points. Far from offsetting recessions, the Federal Reserve has permitted monetary growth to slacken during most postwar recessions, presumably because the pressure comes off interest rates at the trough of the business cycle. Government expenditures have been used only very cautiously to moderate recessions, in spite of the purported commitment of most of the postwar presidents to Keynesian principles. If the multiplier is 0.8 (a reasonable conclusion from reduced-form evidence), and a typical recession involves a shortfall of $60 billion 1979 dollars, then the necessary expansion of the government's use of goods and services is $75 billion, several times larger than anything proposed, much less enacted, in response to a recession.

A second interpretation of the findings is more relevant in the light of the generally noninterventionist policies actually adopted by the United States in the face of recent recessions. The decline in output in the full-employment analysis can also be considered as a decline in potential output of an economy operating at less than full employment. The decline is then interpreted as the cost of permitting a recession to occur as against either a vigorous pro-investment offsetting response or any other way of keeping the economy on a smooth growth path. A deep recession reduces potential output three years later by something over 2 percent, according to this interpretation (a deep recession is one that depresses investment to 10 percent of potential *GNP*). Further, according to the line of argument given above, there is a sense that a recession contributes to subsequent inflation—for the same money stock, the reduction in potential output raises prices.

III. Concluding Remarks and Cautionary Notes

The variations in the rate of capital formation that are associated with *U.S.* business cycles can affect the capital stock by around 7 percent. Correspondingly, full-employment output will vary by a little over 2 percent. For a given money stock, likely differences in price levels on account of the fluctuations are likewise around 2 percent. Though these numbers are fairly small relative to the levels of the corresponding variables, they could be quite important relative to the normal annual changes in the variables. In particular, 2 percent of output over a three-year period is about a quarter of normal growth in output. In this sense, there is an important interaction between the business cycle and the process of capital formation.

Up to this point I have avoided dealing with the welfare implications of the recessionary interruptions in capital formation that occur at random intervals and the related welfare implications of countercyclical policies that try to stabilize investment or output. Most economists will probably take it for granted that more investment is better, so recessions are undesirable, anti-investment policies are even more undesirable, and pro-investment policies are the way we should deal with recessions. There is one justification for this position that will command wide support—the earnings of capital are heavily taxed in the *U.S.* economy, so the social returns to capital formation substantially exceed the private returns, and any trick the government can use to encourage capital formation is to the good. I feel reasonably confident about another conclusion—large negative surprises in monetary policy are undesirable because of their interference with capital formation. But economists are virtually unanimous in their condemnation of sudden reversals in monetary policy, so there is nothing very novel in this conclusion.

The troublesome problem is what to do about an interruption in investment brought about by a nonmonetary shock. From time to time, businesses decide in unison that they need a lower level of inventories, and a classical inventory recession occurs. Further, two sudden increases in world oil prices seem to have contributed to recessions, one in 1974–75 and one apparently about to occur. It is not completely clear to me that fully offsetting the resulting pauses in capital formation is the optimal policy. When people decide to invest less in certain categories, the economy has a variety of ways to make use of the resources. They can be put to use producing other investment

goods (the essential idea of a pro-investment policy), they can be put in the hands of the government (an anti-investment policy), they can produce consumption goods, or they can withdraw from the market economy. Withdrawal seems to be an important part of what actually happens in the United States—recessions are periods of diminished employment and capacity utilization in durable goods industries with some movement of workers from durables to other sectors, but with a substantial increase in time spent away from work by those normally employed in durables. Whether the rather large amount of withdrawal that actually occurs in recessions is efficient is a question we have not really answered yet. Some caution in the use of vigorous pro-investment policies is in order until we understand more clearly exactly what is going on during a recession.

REFERENCES

R. J. Barro, "Unanticipated Money Growth and Unemployment in the United States," *Amer. Econ. Rev.*, Mar. 1977, *67*, 101–15.

______, "Unanticipated Money, Output, and the Price Level in the United States," *J. Polit. Econ.*, Aug. 1978, *86*, 549–80.

S. M. Goldfeld and A. S. Blinder, "Some Implications of Endogenous Stabilization Policy," *Brookings Papers*, Washington 1972, *3*, 585–640.

G. L. Perry, "Potential Output and Productivity," *Brookings Papers*, Washington 1977, *1*, 11–47.

R. M. Solow, "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.*, Feb. 1956, *70*, 65–94.

Issues in the Taxation of Capital Income in the United States

By MICHAEL J. BOSKIN AND JOHN B. SHOVEN*

Few issues in public finance stir as much academic debate or public discussion as the appropriate taxation of income from capital. Prescriptions and practices in various countries range from heavier taxation of capital than labor income to subsidization of investment. A variety of features of the current *U.S.* tax system differentially tax capital income relative to labor income, and certain types of capital income relative to other types. Important examples include the separate corporate income tax, the maximum tax on earned income, the treatment of pensions and life insurance, and the nontaxation of imputed income.

In recent years much insight has been gained into several important issues in analyzing capital income taxation. Applications of the theory of optimal taxation have helped clarify the determinants of the optimal taxation of capital income, new empirical results have questioned long-held beliefs about the effects of capital income taxation on private saving and investment, and careful study has elucidated the complex nature of the incentives involved in the various special features of our current tax laws.

This paper presents a summary of some recent results in this area of public finance. Toward this end, Section I presents a heuristic discussion of the application of optimal tax theory to the taxation of capital income. We note that efficiency may require something other than a convex combination of income and consumption taxation. Optimality might imply heavier taxation of capital income than in an income tax, on the one hand, or an interest income subsidy, on the other. The key parameters in answering this question are the own- and cross-compensated current and forward price effects on consumption and leisure at different stages of the life cycle.

Section II discusses a variety of features of the current *U.S.* tax system which deal with capital income. Of particular interest are the separate corporate tax, the lack of price level indexing, the deductability of interest payments and the marginal finance decision, and the opportunity to save in some forms free of personal taxes. We conclude that the *U.S.* tax system, while often taxing capital income quite heavily, is in many respects a hybrid of an income and expenditure tax.

Section III reviews some recent empirical research which suggests that private saving may be somewhat more responsive to the real after-tax rate of return than previously conjectured, and Section IV offers a brief conclusion and discusses (in light of the analysis in the paper) some policy issues and options currently under serious consideration.

I. Optimal Taxation of Capital Income

Consider first the simplest case in the analysis of the optimal taxation of capital income. Suppose individuals have identical preferences and ability to produce market income. Redistribution is thus ruled out for the moment and the government's sole objective is to raise revenues for public goods. Suppose further that the decision as to the required revenue has been made, perhaps even chosen optimally.[1] Individuals live for two periods (for simplicity, we ignore the different lengths of the periods and assume

*National Bureau of Economic Research and Stanford University. We would like to thank Charles Ballard for his valuable research assistance. We gratefully acknowledge the financial support of the U.S. Treasury.

[1]See Lawrence Lau, Eytan Sheshinski, and Joseph Stiglitz for a discussion of the separability of revenue and spending decisions.

no population growth); in the first they work full time and consume goods, and in the second, labeled retirement, they consume goods and leisure. Thus, their utility may be represented by the usual well-behaved utility function

$$(1) \qquad U = U(C_W, C_R, L_R)$$

where C_W is the annual flow of consumption during work years, C_R is the annual flow of consumption during retirement, and L_R is the annual equivalent of retirement leisure.

The government raises a revenue $\bar{R}$, by taxing consumption in the two periods at tax rates t_W and t_R. A pure consumption tax implies $t_R = t_W$; $t_W > t_R$ implies an interest income subsidy, and $t_W < t_R$ implies an interest income tax. We ignore bequests for the moment.

The government's revenue is then given (in per capita terms) by

$$(2) \qquad t_W C_W + t_R C_R = \bar{R}$$

Individuals seek to maximize utility by choosing C_W, C_R, and L_R subject to their lifetime wealth constraint:

$$(3) \qquad C_W + P_R^f C_R + wL_R \leqslant Y$$

where Y is "full income," w is the wage rate and P_R^f is the forward price of retirement consumption. Work period consumption is taken as numeraire, so $P_{CW} = 1$. In general, we can write the optimal expenditure functions as

$$(4) \qquad \begin{aligned} C_W^* &= g(Y, w, P_R^f) \\ P_R^f C_R^* &= h(Y, w, P_R^f) \\ wL_R^* &= j(Y, w, P_R^f) \end{aligned}$$

Now $P_R^f = P_{CW} e^{-rT}$ where r is the real net return and T is the length of time between saving and dissaving (say, twenty years). Thus,

$$(5) \qquad r = i(1 - (t_R - t_W)) - \pi$$

where i is the nominal before tax rate of return, and π the expected inflation rate. As noted, when $t_W > t_R$, interest is subsidized. We shall ignore inflation in this section, but return to it in Section II.

Solving the usual constrained maximization problem to minimize the deadweight loss from t_W and t_R, or to maximize utility subject to raising the required revenue yields

$$(6) \qquad \frac{t_w^*}{t_R^*} = \frac{(C_w^2/P_R^f + C_R)\eta_{C_W, C_R} + C_W \cdot C_R \cdot \eta_{C_R, L_R}}{(C_w^2/P_R^f + C_R)\eta_{C_W, C_R} + C_W \cdot C_R \cdot \eta_{C_W, L_R}}$$

where η_{ij} is the compensated elasticity of i with respect to j (ranging over the three goods discussed above). Thus, a pure consumption tax is optimal *only* when working period and retirement consumption are equally good substitutes for retirement leisure.[2] The optimal tax system in this stylized world taxes the weaker substitute for leisure during retirement more heavily than the closer substitute. Thus, if working-period consumption and leisure late in life are weaker substitutes than are retirement consumption and leisure, an interest income subsidy would be desirable to achieve second best efficiency.

In addition to the stylized assumptions discussed above, we have also implicitly assumed a no-growth economy: a series of overlapping generations of equal size and no technical progress. In a growing economy, we need to append a "Golden Rule" (see Edmund Phelps) condition to the Ramsey rules for optimal taxation (as discussed by M. King). A fuller set of government instruments, for example, monetary policy, government investment, etc. could be used to influence saving and investment, but we ignore such elaborations here.

We recognize that individuals do not have identical abilities to provide income, nor identical preferences, nor identical *ex post* outcomes from identical *ex ante* possibilities

[2] As derived in many places, for example, A. B. Atkinson and Stiglitz, Martin Feldstein (1978), Boskin and Lau, among others.

in an uncertain world. The "ability" to generate income differs because of differences in underlying attributes such as intelligence, initial inheritance of capital, early parental investment in human capital, etc. Such complexities have been dealt with in several separate papers,[3] but never comprehensively. Heuristically, differences in income due to underlying attributes can be dealt with via credits and/or a progressive rate structure in a personal tax system which meets the optimal tax criteria described above.

If much of the difference in lifetime inome were due to differential inheritance —a source of much debate— the joint density of "innate ability," or wages, and inheritance would be relevant in the optimal taxation of capital income. A high positive correlation between ability and inheritance might lead to heavy taxation of bequests for distributional reasons; of course, such taxation would have its own incentive effects (see Stiglitz, 1978) which would have to be considered. The problem of bequests is particularly acute because of the controversy over how much saving is for life-cycle purposes and how much due to planned bequests. (See L. J. Kotlikoff and L. Summers, and Boskin, 1977.) We have discussed the optimal taxation of capital income in a simple life cycle context; an overwhelming bequest motive in saving adds substantial complexity to the optimal tax formulae.

With these very brief comments on the optimal taxation of capital income in mind, let us turn to a discussion of the current *U.S.* tax system and its treatment of income from capital.

II. Current *U.S.* Tax Treatment of Capital Income

The *U.S.* tax system is very complex and does not even vaguely represent a true Haig-Simons income tax. Capital income in particular is subject to a vast array of separate treatments, the combined effect of which is difficult to assess. First, some of capital income, namely corporate equity income, is taxed twice, first at the corporate level and second at the personal level. However, at the corporate level interest payments are tax deductible, and therefore the distortionary effects of the corporate income tax depend on the financial composition of marginal investment (see Stiglitz, 1976, for example). At the personal level, dividends are fully taxable except for the $100 exclusion, while capital gains are only 40 percent taxable, and even that tax is delayed until realization. However, capital gains are defined in nominal terms. Pure inflation turns the capital gains tax into a wealth tax on equity holders, the rate of which depends on the rate of inflation, and the collection of which occurs upon realization (see Feldstein and J. Slemrod). Clearly, high rates of inflation increase the total tax borne by the owners of capital and add new strength to the "lock-in effect." Essentially, with inflation a substantial transactions tax is imposed on those who reallocate their investments. Inflation also increases the total taxation of the earnings of capital in the corporate sector by reducing the real present value of depreciation deductions which are based on historical costs.

The heavy taxation of corporate equity income, exacerbated by inflation, does not imply that there are no vehicles for saving which are lightly taxed. Approximately 50 percent of savings is currently taxed on an expenditure tax basis at the personal level. All consumer durables and owner-occupied housing are purchased with after-tax dollars, but the imputed return on these items is not taxed. That is, the accumulation and return on these assets is taxed once and not twice as implied by an income tax. Taxing the acquisition rather than the return flow is referred to (and advocated for these items) as the "prepayment" method in the Treasury Department's *Blueprints for Basic Tax Reform*. To the extent the purchase price is equal to the present value of future returns, taxing the acquisition is equivalent in present value terms to taxing the imputed earnings.[4] According to the *Flow of Funds*

[3]See Atkinson and Stiglitz, Stiglitz (1978), and Boskin (1980).

[4]Houses purchased outright are presently treated on the "prepayment" expenditure tax basis. However, mortgage purchases would be treated differently under

Accounts, roughly 20 percent of net savings are made in the form of net accumulation of new owner-occupied housing.

A significant amount of savings also flows through state, local, or federal government pension plans (excluding Social Security), through private pension plans, and through cash value life insurance policies. These are either taxed on a deferral basis (where both the initial contribution and the return are sheltered from taxation but the payouts are fully taxable) or on a prepayment basis (where the initial contribution is subject to tax but both the return and payout are tax free). It is simple to show that the two methods are identical except that the tax rate at the time of withdrawal may be different from the rate at the time of contribution. Again, *Flow of Funds* indicate that in recent years approximately 30 percent of savings flow through these vehicles and are thus taxed on a consumption tax basis at the personal level (the underlying assets may still be subject to the corporate income tax).

Although one-half of private saving is done in tax sheltered forms, the heavy taxation of the other half of saving may have reduced the real net return to all saving. Put simply, equalization of after-tax rates of return net of risk has created a situation where even those types of savings which are not directly taxed twice may be earning a net return less than they would earn under a comprehensive expenditure tax. We shall return to an important implication of this point in Section IV.

With capital income from some sources taxed more heavily than labor income, while that from other sources (or held via particular vehicles) is taxed very lightly, the current tax system distorts the allocation of investment far more than necessary. This distortion across industrial allocation and across holding forms may, indeed, be so large as to offset any apparent advantages of partial or piecemeal movement towards an expenditure tax.

a true expenditure tax. Whereas now interest payments are deductible, they would not be under an expenditure tax.

III. Recent Empirical Evidence

The relative constancy of the gross private saving rate—the ratio of gross private saving to gross national income—so well documented by Paul David and John Scadding fails to reveal a variety of important features of private saving in the United States. The gross private saving rate is the product of the saving rate out of disposable income and the ratio of disposable income to total income, i.e.,

$$(7) \quad GPSR = \frac{GPS}{GNP} = \frac{GPS}{DPI} \times \frac{DPI}{GNP}$$

We know that taxes as a percentage of total income have risen substantially over this period. Hence, the saving rate out of disposable income must have increased substantially to offset the decline in the ratio of private to total income. The behavioral interpretation given to these data by David and Scadding is that taxes and present consumption are essentially perfect substitutes: the rise in taxes is offset by an equivalent decline in current consumption. They go on to explore a variety of intriguing conjectures concerning consumer behavior.

Three basic points need to be made in criticism of this conjecture. First, most theories of consumer behavior relate saving to disposable income. If this is correct, the saving rate varies substantially. Second, it indeed would be surprising if consumers made this type of rational calculation vis-à-vis the government and business sectors in terms of gross saving and income. Consumers know their capital depreciates. Again, our economic theories generally relate to how consumers choose their net position. Further, except for some possible embodied technical change, it is net saving that is relevant to the issue of whether taxes affect capital accumulation. The net private saving rate—net saving divided by net income—exhibits substantially more relative variation than the gross series and can hardly be called constant, even if we confine ourselves to the postwar period.[5] While de-

[5]If one took the broader view of saving as inclusive of human investment, use of John Kendrick's data reveals still more variability in the total saving rate, gross as well as net.

preciation series are notoriously unreliable, use of several alternative depreciation series (for example, tax, replacement cost, etc.) yields substantial variation in the net private saving rate. We take this to be a strong indictment of the structural interpretation of Denison's Law. Third, even if total gross income and gross saving are examined, there still may be an independent effect of real net rates of return on saving. Even if taxes and present consumption are perfect substitutes (the public sector is doing its benefit-cost analyses properly, free-rider issues are ignored, etc.), the share of private wealth consumed today (publicly or privately) will depend upon the net, or after-tax, return to saving, whereas gross income is the flow from private wealth at the gross return. Hence, taxes decreasing the net return to saving may cause a decrease in saving.

For several decades, econometric work on saving behavior consisted largely of estimating Keynesian-type consumption functions. The inclusion of an interest-rate variable in such analysis was the exception rather than the rule. Further, when interest rates were included, nominal before-tax rates rather than real after-tax rates were used. Feldstein (1970) has demonstrated that such a procedure almost certainly biases downward the estimated interest elasticity. Since most of the early work on consumption and saving focused on issues other than the effect of interest rates, perhaps it is not surprising that little attention was paid to the weak, and sometimes negative, relationship between saving and the rate of interest. Richard Musgrave and Peggy Musgrave report that "studies of the relationship between saving and the rate of interest differ in their conclusion. Some hold that there is a substantial negative relationship, while others attribute little weight to the rate of interest in the consumption function" (p. 478). It is curious, however, that little attention is paid to interest rates in consumption functions in the large-scale econometric macro models in widespread use today.

Several recent studies of saving have included interest rates as determinants of saving. Colin Wright includes a measure of after-tax rates of return on stocks and bonds in estimating consumption functions from *U.S.* annual time-series data. His estimates imply an interest elasticity of saving of approximately 0.2. As he notes, this is substantially larger than the usual assumption and, despite his efforts, may be closer to the total than the pure-substitution elasticity. However, his measures of consumption and income suffer from several deficiencies, and his data refer to the period prior to 1958. Hence, at the very least, his results must be improved and updated.

Warren Weber (1970, 1975) examines the impact of interest rates on aggregate consumption. He finds a positive relationship between consumer expenditures and nominal interest rates. In the second study, he includes the expected inflation rate as a determinant of consumer expenditures but finds no evidence that expected inflation affects consumption.

In a study of quarterly *U.S.* aggregate postwar data, L. Taylor estimates an enormous interest elasticity, approximately 0.8. Since his study is directed toward other issues, he merely reports this result without attempting to explain why his estimate is several times larger than those of other researchers. Perhaps this is because it is unclear that he is estimating a structural equation rather than a reduced form from some larger system.

Finally, in a thought-provoking reexamination of Denison's Law, David and Scadding document the continued constancy of the gross private saving rate, the constancy of the saving rate augmented to include consumer durables purchases in saving and the rental flow from durables in income, and changes in the composition of private saving between the household and business sectors. They interpret this relative constancy of the gross private saving rate as evidence that taxes—either through a reduction in private income or a reduction in the real net rate of return on capital—do not affect private saving behavior. While this argument also has been made by a large number of other economists, recent work questions drawing such behavioral inferences from these data.

Recently, Boskin (1978) and Boskin and Lau have reexamined econometrically the effects of real net rates of return on the consumption/saving choice. Boskin attempts to get much better measures of the relevant variables—consumption rather than consumer expenditure; expected long-run real net rates of return rather than nominal or gross or short-run rates, etc. —and uses somewhat more sophisticated econometric procedures to estimate consumption functions. Using a variety of specifications of the consumption function, sample periods, and measures of the relevant variables, he concludes that the real net rate-of-return elasticity of private saving is about 0.4 per annum. In life cycle models which include the indirect effect of interest rates on wealth, King and Summers derive overall interest elasticities of 1.5 to 2.5. These estimates are consistent with those in Boskin and Lau where an explicit translog utility function is used to derive functional forms for current consumption and labor supply functions.

While these findings need to be supplemented by further studies before extreme confidence can be placed in the exact estimates, they clearly suggest that heavy taxation of capital income substantially curtails private saving by reducing the real net rate of return and that restructuring the taxation of income from capital in the United States may well be the top priority for tax reform in the next decade. We now turn to some such issues.

IV. Policy Evaluation and Conclusion

While it is clear that further investigation of the magnitude of the key elasticities is needed, policy analyses making particular assumptions regarding these variables are beginning to appear. In a recent paper, Donald Fullerton, Shoven, and John Whalley estimated the efficiency gains which the economy could realize by removing the corporate income tax and by making all saving deductible at the personal level. They assumed the long-run elasticity of savings with respect to the real after-tax rate of return was 0.4. The results indicated that the present value efficiency gain would be between \$200 and \$600 billion (1973 dollars), depending on which tax vehicle is used to maintain government revenue. Interestingly, they found that removing the corporate income tax, but making all saving taxable (i.e., establishing a true income tax) also offered an efficiency improvement, this time between \$165 and \$220 billion. The conclusion was that while the expenditure tax would improve economic efficiency, and while only half of net savings are taxed under the current system in the United States, virtually no efficiency improvement has been realized by this policy of allowing some tax shelters for saving. The authors estimated that the entire intertemporal efficiency gain was eliminated by the distortions which were created in the interindustry allocation of capital.

Summers has also examined the efficiency gain offered by an expenditure tax. His estimates are far larger than those of Fullerton, Shoven, and Whalley primarily because he uses a long-run elasticity of 2.0 and uses a lower discount rate in his present value efficiency evaluation.

We anticipate that the general result that moving towards an expenditure tax offers *potential* gains in allocational efficiency will prove robust to other studies. The lesson of this paper, however, is that increased incentives for saving should not be available only for certain narrow industries or forms of saving. If this policy is continued, then the potential gains are likely to be sacrified. Perhaps the most attractive route towards an expenditure tax would involve an extreme liberalization of Keogh and IRA pension-savings vehicles. If the limits on these were raised, if the assets which could be held were very inclusive, and if the taxable withdrawal could be made at any time, these institutions could effectively institute an expenditure tax.

In conclusion, we want to restate what has been accomplished by the profession in this area and what needs further work. It seems that the key determinants of the optimal taxation of capital income have been identified. Some progress has been made in

estimating these elasticities, with the general result that they are larger than had been previously assumed. Some progress has been made toward evaluating the magnitude of the efficiency gain offered by moving towards an expenditure tax, but additional work is needed, both in estimating the key elasticity parameters and in evaluating tax policy revisions.

REFERENCES

A. B. Atkinson and J. E. Stiglitz, "The Design of Tax Structure: Direct versus Indirect Taxation," *J. Publ. Econ.*, July/Aug. 1976, *6*, 55–75.

M. J. Boskin, "Interrelationships between the Choice of Tax Bases, Tax Rates and the Unit of Account in the Design of An Optimal Tax System," in Henry Aaron and Michael Boskin, eds., *Studies in the Economics of Taxation: Essays in Honor of Joseph Peckman*, Washington 1980.

——, "Taxation, Saving and the Rate of Interest," *J. Polit. Econ.*, Apr. 1978, *86*, S3–S27.

——, "Is Heavy Taxation of Capital Socially Desirable?" in *The Role of Federal Tax Policy in Stimulating Capital Formation and Economic Growth*, Joint Economic Committee Hearings, 95th Cong., 1st sess., July 1977.

—— **and L. J. Lau,** *Taxation, Social Security and Aggregate Factor Supply in the United States*, Washington 1978.

P. David and J. L. Scadding, "Private Saving: Ultrarationality, Aggregation, and 'Denison's Law'," *J. Polit. Econ.*, Mar./Apr. 1974, *82*, 225–49.

M. S. Feldstein, "The Welfare Cost of Capital Income Taxation," *J. Polit. Econ.*, Apr. 1978, *86*, S29–S51.

——, "Inflation, Specification Bias, and the Impact of Interest Savings Rates," *J. Polit. Econ.*, Nov./Dec. 1970, *78*, 1325–39.

—— **and J. Slemrod,** "Inflation and the Excess Taxation of Capital Gains on Corporate Stock," *Nat. Tax J.*, June 1978, *31*, 107–18.

D. Fullerton, J. B. Shoven, and J. Whalley, "Dynamic General Equilibrium Impacts of Replacing the U.S. Income Tax with a Progressive Consumption Tax," Nat. Bur. Econ. Res. work. paper, 1979.

John Kendrick, *The Formation and Stocks of Total Capital*, New York 1979

M. King, "Savings and Taxation," in G. A. Huges and G. M. Heal, eds., *Essays in Public Policy*, London 1980.

L. J. Kotlikoff and L. Summers, "The Role of Intergenerational Transfers in Aggregate Capital Accumulation," unpublished manuscript, Univ. California-Los Angeles 1979.

L. J. Lau, E. Sheshinski, and J. E. Stiglitz, "Efficiency in the Optium Supply of Public Goods," *Econometrica*, Jan. 1978, *46*, 269–84.

Richard A. Musgrave and Peggy B. Musgrave, *Public Finance in Theory and Practice*, New York 1973.

E. Phelps, "Second Essay on the Golden Rule of Accumulation," *Amer. Econ. Rev.*, Sept. 1965, *55*, 793–814.

F. P. Ramsey, "A Contribution to the Theory of Taxation," *Econ. J.*, Mar. 1927, *37*, 47–61.

J. E. Stiglitz, "Notes on Estate Taxes, Redistribution and the Concept of Balanced Growth Path Incidence," *J. Polit. Econ.*, Apr. 1978, *86*, S137–150.

J. E. Stiglitz, "The Corporation Tax," *J. Publ. Econ.*, Apr./May 1976, *5*, 303–11.

L. H. Summers, "Tax Policy in a Life Cycle Model," Nat. Bur. Econ. Res. work. paper no. 302, 1978.

L. Taylor, "Saving out of Different Types of Income," *Brookings Papers*, Washington 1971, *2*, 383–407.

W. E. Weber, "Interest Rates, Inflation and Consumer Expenditures," *Amer. Econ. Rev.*, Dec. 1975, *65*, 843–71.

——, "The Effect of Interest Rates on Aggregate Consumption," *Amer. Econ. Rev.*, Sept. 1970, *60*, 591–600.

C. Wright, "Saving and the Rate of Interest," In Arnold C. Harberger and Martin J. Bailey, eds., *The Taxation of Income from Capital*, Washington 1979.

Board of Governors of the Federal Reserve System, *Flow of Funds Accounts*, 3d quarter, Nov. 1977.

U.S. Treasury, *Blueprints for Basic Tax Reform*, Washington 1977.

What Is Labor Supply and Do Taxes Affect It?

By HARVEY S. ROSEN*

The issue of tax-induced changes in labor supply behavior has been receiving increasing attention. Economic theory alone can say little about the impact of income taxation on labor supply because of the well-known conflict between income and substitution effects. Therefore, an enormous amount of effort has been devoted to empirical investigation of this problem, with a focus on the impact of taxes on hours of work and labor force participation rates. In Section I of this paper, I briefly discuss this literature and its major conclusions.

It has been long understood, however, that the concept "labor supply" is more general than "hours of work." If one individual is healthier, better educated, and more highly motivated than another, then presumably a given number of hours of work will lead to a greater effective labor supply for the former than for the latter. Thus, studies of the effect of taxes on other dimensions of labor supply are needed in order to assess the full impact of taxes on work incentives. The main purpose of this paper is to discuss some of this research (Section II) and to explore its policy implications (Section III).

I. Taxes and Hours of Work

Econometric studies of taxation and hours of work have increased steadily in sophistication. In the pioneering econometric studies of the determinants of hours of work, taxes were mostly ignored. Gross rather than net wages were included in the list of regressors. (See, for example, Glen Cain, or William Bowen and T. Aldrich Finegan.) A major advance was due to Robert Hall, who recognized that a progressive tax system fundamentally changes the shape of the budget constraint in income-leisure space (it is kinked, rather than a straight line), and therefore, standard econometric techniques must be modified. Subsequent work has dealt more explicitly with tax-induced nonlinearities; the reader is referred to Gary Burtless and Jerry Hausman who discuss the problem in the context of the negative income tax.

The more econometrically advanced studies have tended to confirm the substantive results of their predecessors. Although considerable differences inevitably arise in parameter estimates due to differences in samples, time periods, and statistical techniques, I think it would be fair to say that two important "stylized facts" have been isolated:

1. For prime age males, the substitution effect of changes in the net wage upon hours of work tends to be small in absolute value and often statistically insignificant. This result has emerged from both cross-sectional studies (see, for example, Cain and Harold Watts, or Burtless and Hausman) and from time-series studies (see, for example, Robert Lucas and Leonard Rapping).[1]

2. The hours of work and labor force participation decisions of married women are quite sensitive to changes in the net wage. Although estimates differ widely, a number of investigators have found elasticities in excess of 1.0 (see Cain and Watts, the author, or Reuben Gronau).

*Assistant professor, Princeton University, and research associate, National Bureau of Economic Research. This research was supported by NSF Grant No. SOC-7907847. I would like to thank C. Reimers, P. Hendershott, J. Eaton, and J. Pechman for useful suggestions.

[1]An implication of this research is that in the short run, at least, it is unlikely that reductions in labor's tax burden will unleash an enormous amount of work effort.

One can imagine these results leading to some complacency on the part of individuals who conduct tax policy because of the implication that, except for secondary workers, increased taxation will not have much of an impact on hours of work decisions.[2] The discussion in the next section suggests that caution may nevertheless be appropriate.

II. Taxes and Other Dimensions of Labor Supply

The literature's emphasis on hours of work is easy to understand because it is an important variable and one that is relatively straightforward to measure. Nevertheless, labor supply is a concept more general than number of hours per week or per year. In particular, at least three important dimensions of labor supply, other than hours, may be influenced by taxes: (a) lifetime hours of work and timing of retirement; (b) intensity of work effort; and (c) quality of work effort.

The theoretical and empirical evidence on these important issues is currently rather scanty. In this section, I discuss and evaluate some fragmentary evidence and suggest a few possibilities for future research.

A. *Time of Retirement*

Although income taxes[3] may not influence the number of hours worked per year, they may affect the number of years worked, that is, the retirement decision. As Michael Boskin (1977) has noted, until fairly recently it was widely believed that financial variables had little to do with retirement decisions. The prevailing opinion, based upon survey interviews, was that bad health was the overriding reason for retirement. In the past few years, however, a number of econometric studies of the determinants of the retirement decision have appeared (see Boskin, 1977; Richard Burkhauser; Alan Blinder et al.). Almost uniformly, these studies suggest that the availability of pensions, the implicit tax imposed on wages by the Social Security system, and other financial considerations have statistically significant and quantitatively important impacts on the probability of retirement.

Most of these studies have focused on the interaction between pension and social insurance systems, on the one hand, and the retirement decision, on the other. Nevertheless, in some cases the results can be used to shed some light on the question of how the personal income tax influences retirement decisions, *ceteris paribus*. This is possible because the explanatory variables generally include potential market earnings, which depend upon the tax rate.

For example, Burkhauser analyzes the probability of males accepting Social Security retirement benefits at age 62 as a function of marital status, education, potential market earnings, and several other variables.[4] His estimates suggest substantial responsiveness of retirement probability with respect to market earnings (p. 19). Consider, for example, an individual with adjusted gross income (in 1973) of $9,700. Assuming that he pays a payroll tax of 6.05 percent and an average income tax of 9.0 percent,[5] his disposable income would be $8,240. According to Burkhauser's figures, this $1,460 change in net income would induce an increase in excess of 4 percentage points in the probability of retirement at age 62, a substantial difference.

This computation is meant to be only illustrative. Personal income taxes have not yet been integrated with sufficient care into models of the retirement decision to allow more definitive conclusions. However, in light of the apparent sensitivity of retirement decisions to economic variables, this is a subject worth further inquiry.

[2]Of course, due to income effects, there can be a substantial excess burden of taxation, even if the uncompensated labor supply elasticity is small.

[3]In this paper I consider only taxes defined in the narrow sense, and not benefit reductions in transfer programs.

[4]Burkhauser's data were from the 1973 Social Security Exact Match File.

[5]This number is for a married person with no dependents. See Tax Foundation Inc., p. 105.

B. *Intensity of Work*

"Hours of work" are usually measured as the amount of time elapsed at the workplace. However, because intensity of effort varies from individual to individual, elapsed hours may be a poor proxy for the effective number of hours of work. Clearly, intensity of effort is difficult to measure, particularly in jobs where salaries are not based on piece rates. An interesting attack on this measurement problem has recently been made by Frank Stafford and Greg Duncan who analyze data from the *Time Use Survey* administered by the Survey Research Center of the University of Michigan. These data contain detailed time diaries in which individuals report on several measures of on-the-job time allocation: time spent in formal or scheduled work breaks; time spent informally socializing or any other type of unscheduled work breaks; and time spent in on-the-job training.

With this information, Stafford and Duncan can compute an estimate of effective number of hours, as well as hours of work as conventionally measured. In an interesting experiment, they estimate two labor supply equations for men, one using conventional hours as the dependent variable, and the second with effective hours. The right-hand side variables include education and age. The coefficients on education and age in the second equation have larger (positive) signs and relatively smaller standard errors than their counterparts in the first (p. 34). To the extent that education and age are proxying for the wage, these results suggest that improper measurement of effective hours may be obscuring a positive wage response.

Because Stafford and Duncan do not explicitly include the after-tax wage in their effective hours equation, nothing more specific can be said at this time. But the use of time diaries seems a fruitful way to continue work on this problem.

C. *Human Capital Investments*

It has been argued that one of the key sources of wealth in the United States and other developed countries is human capital, which augments the productivity of workers, and thereby increases the effective size of the labor force. Despite intensive analysis of human capital decisions,[6] the relation of taxes to human capital investment has received little attention.

Schultz has argued that the *U.S.* tax system discriminates against human capital investment: "Our tax laws everywhere discriminate against human capital. Although the stock of such capital has become large and even though it is obvious that human capital, like other forms of reproducible capital, depreciates, becomes obsolete, and entails maintenance, our tax laws are all but blind on these matters" (p. 13).

This view has been challenged by Boskin, who points out that Schultz's argument regarding the lack of deductibility for expenditures on human capital investment is erroneous once it is recognized that the most important costs of human capital are foregone earnings, rather than tuition payments. It has been estimated that foregone earnings amount to over half the costs of human capital (see Boskin, 1975, p. 5).

Boskin points out that if all the costs of human capital investment are foregone earnings, then in a simple model a proportional wage tax has *no* impact whatsoever on the decision to invest in human capital. The logic of this argument is straightforward. The tax reduces the benefits and costs in the same proportion, so if the net present value was positive prior to the tax, it remains positive after the tax is imposed.

However, Jonathan Eaton and I have shown that even under the assumption that the only costs of human capital accumulation are foregone earnings, the neutrality result does not necessarily follow. There are two reasons for this. First, an important assumption in the simple model is that the supply of hours of work is fixed regardless of the net wage. Now, hours of work can be thought of as the "utilization rate" of human capital: the more the individual works, the higher is the rate of utilization and,

[6]See, for example, Theodore Schultz or Gary Becker.

therefore, the higher the return on the human capital investment. With an endogenous hours of work decision, then, a wage tax may change the benefits of the investment without an offsetting movement in the costs, and neutrality no longer holds.

A second factor ignored in the simple model is the uncertainty of returns to human capital. When an individual makes an educational investment, he or she does not know for certain that it will increase earnings capacity, or by how much.[7] It can be shown that even when hours of work are independent of the net wage, if the returns to human capital are stochastic, then proportional wage taxation will not in general be neutral in the human capital decision (see Eaton and myself). Rather, the impact of taxation is ambiguous because of two conflicting effects: (i) A proportional wage tax cuts the riskiness of human capital because, in effect, the Treasury serves as the taxpayer's silent partner, sharing in both gains and losses.[8] To the extent that the individual is risk averse, this insurance effect tends to *increase* human capital accumulation. (ii) On the other hand, the proportional tax reduces the individual's wealth. To the extent that the desire to invest in relatively risky assets decrease with wealth, then this effect will tend to *decrease* investment in human capital.

Because the two effects work in opposite directions, in the absence of specific assumptions on how risk aversion varies with wealth, it is impossible to know a priori how a proportional wage tax will change human capital accumulation. Just as in the hours of work case, only empirical work can settle the question. However, measurement problems make econometric analysis here even harder than in the hours of work case. How does one measure the amount of capital embodied in a human being? What proportion of educational expenditures are consumption and what proportion investment? How can one estimate the amount of earnings foregone in on-the-job and vocational training programs?

In light of these formidable problems, it is no wonder that little empirical work has been done. However, a recent paper by Robert Willis and Sherwin Rosen permits some speculation as to the impact of taxes on an important kind of human capital investment, college attendance.

The Willis-Rosen (hereafter W-R) study measures the influence of various factors on the probability that an individual attends college. They investigate a number of "background" variables, such as religion, achievement test scores, and father's occupation. In addition, they find that the probability is significantly affected by the expected growth rates of earnings, with and without college, and by the ratio of the initial earnings obtained by college graduates to the earnings of those who did not attend.

Some of their results are reported in Table 1. The variable g_c is the expected growth rate of earnings if the individual goes to college, g_{nc} is the rate without college, and y_c^i/y_{nc}^i is the ratio of college to noncollege earnings in the initial period (the elasticities are evaluated at a sample proportion of one-half). These figures suggest that returns to college exert a substantial impact upon the decision to attend.

An interesting exercise is to compute the percentage changes in g_c, g_{nc}, and y_c/y_{nc} generated by the federal income tax, and use the elasticities in Table 1 to find the implied change in the probability of enrolling in college. More specifically, for pretax values of the variables, I use the means reported by W-R (p. 41). Posttax magnitudes are calculated by applying federal tax tables to the pretax values.[9] Due to progressivity, the income tax lowers g_c by 10.1 percent, g_{nc} by 12.2 percent, and y_c/y_{nc} by 4.6 percent. Taking these results together

[7]It is unlikely that the individual can insure himself against such risks because the problems of moral hazard associated with insurance in general are especially pervasive in the case of human capital. In such a situation, the market is unlikely to provide insurance.

[8]Similar arguments have been used in discussion of the impact of taxation on nonhuman investment. See, for example, Joseph Stiglitz.

[9]For purpose of computing tax liability, it is assumed that the individual is married, files jointly, and has two dependents.

TABLE 1

Elasticities of the Probability of College Attendance with Respect to:		
g_c	g_{nc}	y_c/y_{nc}
2.91	−0.61	4.1

Source: Based upon estimates in Willis-Rosen and included in a personal communication from Willis.

with the elasticities in Table 1 yields a decrease of 3.5 percentage points in the probability of enrolling in college. Because the elasticities of Table 1 are at the low end (in absolute value) of those reported by W-R, this figure should be regarded as a conservative one.

As Willis and Rosen emphasized, their results cannot be extrapolated to the population as a whole because the sample used to generate the parameter estimates is not representative. (It consists of a group of male World War II veterans who applied for the Army Air Corps.) However, the estimates suggest that enough of an effect may be present to make further empirical investigation of this problem worthwhile.

III. Conclusions

There is a rich and extensive literature on the impact of taxation upon hours of work. Doubtless, econometric work to refine current estimates will continue. But "labor supply" and "hours of work" are not the same, and it is the former concept that should be of concern to those who are interested in the long-run economic effects of taxation. As noted above, this fact has long been realized, but I think that an implicit view in much of the public finance literature is that it is simply too difficult to find out anything useful about the effect of taxes on other dimensions of labor supply. I have tried to show that there exists research which can be built upon to learn about this subject, and even some hints that the tax effects are important.

At this point, all that can be said to the makers of tax policy is that they should proceed with caution when it comes to taxing labor. As evidence accumulates that the income tax and social insurance systems appreciably influence the stock of physical capital (see, for example, Martin Feldstein), there may be a temptation to increase relative tax rates on labor because hours of work appear to be inelastic in supply. But hours of work are just the tip of an iceberg that is potentially very deep, and this may be a misguided policy.

REFERENCES

Gary Becker, *Human Capital,* New York 1964.

A. S. Blinder et al., "Market Wages Reservation Wages, and Retirement Decisions," mimeo., Princeton Univ. 1978.

M. Boskin, "Notes on the Tax Treatment of Human Capital," mimeo., Nat. Bur. Econ. Res., Stanford 1975.

______, "Social Security and Retirement Decision," *Econ. Inquiry,* Jan. 1977, *15,* 1–25.

William G. Bowen and T. Aldrich Finegan, *The Economics of Labor Force Participation,* Princeton 1967.

R. V. Burkhauser, "An Asset Maximization Approach to Early Social Security Acceptance," unpublished paper, Inst. Res. Poverty, Univ. Wisconsin 1977.

G. Burtless and J. A. Hausman, "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *J. Polit. Econ.,* Dec. 1978, *86,* 1103–30.

Glen G. Cain, *Married Women in the Labor Force,* Chicago 1967.

______ **and Harold Watts,** *Income Maintenance and Labor Supply,* Chicago 1973.

J. Eaton and H. S. Rosen, "Taxation, Human Capital, and Uncertainty," Ind. Relat. Sec. paper no. 117, Princeton Univ. 1979.

M. S. Feldstein, "The Welfare Cost of Capital Income Taxation," *J. Polit. Econ.,* Apr. 1978, *86,* 529–52.

R. Gronau, "Leisure, Home Production, and Work—The Theory of the Allocation of Time Revisited," *J. Polit. Econ.,* Dec. 1977, *85,* 1099–123.

R. E. Hall, "Wages, Income and Hours of Work in the U.S. Labor Force," in Glen

Cain and Harold Watts, eds., *Income Maintenance and Labor Supply*, Chicago 1973.

R. Lucas and L. Rapping, "Real Wages, Employment, and Inflation," in Edmund Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

H. S. Rosen, "Taxes in a Labor Supply Model with Joint Wage-Hours Determination," *Econometrica*, May 1976, *44*, 485–507.

T. W. Schultz, "Investment in Human Capital," *Amer. Econ. Rev.*, Mar. 1961, *51*, 1–17.

F. Stafford and G. Duncan, "The Use of Time and Technology by Households in the United States," mimeo., Univ. Michigan 1977.

J. Stiglitz, "The Effects of Income, Wealth and Capital Gains Taxation on Risk Taking," *Quart. J. Econ.*, May 1969, *83*, 263–83.

R. Willis and S. Rosen, "Education and Self-Selection," mimeo., Univ. Chicago 1978.

Tax Foundation Inc., *Facts and Figures on Government Finance*, New York 1977.

Private Saving

By GEORGE M. VON FURSTENBERG*

Many of the difficulties in explaining the behavior of private saving arise from the fact that, as measured in the national income and product accounts (*NIPA*), private saving does not purport to measure the change in the real value of private net worth. Rather, this change in net worth is composed of (deflated) *NIPA* personal saving (*PS*) plus real net asset revaluation gains. Such gains may be due in part to corporate saving (*CS*), defined as undistributed corporate profits with inventory valuation (*IVA*) and capital consumption adjustments. Financial regulations and the capitalization of outdated conventions of our tax laws may cause increases in the rate of inflation to depress real household net worth for a time by more than private saving can raise it. Thus, in spite of $253 billion of net private saving over the 1972–74 period, real household net worth declined by around 25 percent from the end of 1971 to the end of 1974. The volatility of market valuations also appears to rise with the rate of inflation, thereby further reducing the positive covariance between changes in household net worth and *PS*. This makes it less tenable to use theories and explanatory variables of the demand for household net worth as if they applied equally to the supply of personal saving.

Specifically, large variations in actual net worth that are due to an ill-defined mixture of stochastic and nonrandom changes in real values make it hazardous to explain *PS* as the difference between the actual net worth at the beginning and the net worth desired for the end of the accounting period over which saving is measured. Some disturbances will average out over time, others need not. In either event they make it extremely difficult to determine the expected after-tax real rate of return on actual or potential household portfolios that may enter into the determination of desired net worth. At the micro level, the growth of a household's net worth may now depend much more on successful asset, liability, and tax management than on abstention from current consumption. In fact, more management, advisory, and loan capital services may have to be consumed to get ahead. David Hanson and Carmen Menezes have recently reminded us of Alfred Marshall's conjecture that capital risk affects saving adversely. All this casts some doubt on the behavioral identification of any regularities that may be found in the *NIPA* time-series on personal saving.

Little progress has been made in recent research to reduce these doubts although systematic, depressing effects of inflation on the valuation of financial claims to nonfinancial corporations relative to the replacement cost of their assets (James Tobin's q) have recently been identified. The investment characteristics of business equities are increasingly being contrasted with the superhedge characteristics of real estate (and gold) against rising inflation rates. For instance, inflation raises the tax rate on adjusted business profits at the same time as it raises the tax subsidy rate on mortgaged homes. In this as in other respects it changes the variance-covariance matrix of real after-tax rates of return as well as the absolute and relative size of the real rates on different assets and liabilities. It has also changed the weight of different assets in consumer portfolios as William Fellner has shown.

To follow much of the recent literature in assuming that the expected after-tax real rate of return on a single instrument, such as Aaa or Baa corporate bonds, is an equally acceptable proxy for the expected after-tax return on the total household portfolio, no matter how much inflation has

*International Monetary Fund, on leave from Indiana University. I am solely responsible for the views expressed.

changed, would take an act of blind faith. In actuality there need be little or no correlation between the two rates over time. Up to now at least, the expected after-tax real rate of return on household net worth and the probability distribution of the factors thought to condition such expectations have been approximated only arbitrarily and sketchily in the literature. Finding a positive or negative effect on personal saving appears to depend on the choice and processing of questionable input. In addition, the underlying rate of inflation and the real after-tax rates of return constructed by almost any measure move so consistently in opposite directions over time that their effects are difficult to separate even when discrimination is attempted. Policymakers could therefore expect no reliable guidance from empirical analyses of aggregate *U.S.* time-series on the question of whether saving incentives "work," even if there were no doubt that higher saving rates *ex ante* produce higher investment rates *ex post* before long.

Some progress has been made on other fronts to take a more comprehensive view of the determinants of private saving. Chief among these is the recognition that (permanent) disposable personal income does not comprise all the income flows on which personal saving decisions are based. Rather, income flows that are siphoned off by the government or retained by corporations may also contribute to the current or future resources available to households or increase their net worth.

Dealing with the net national product (*NNP*) and all its components rather than just disposable income (*YD*) creates certain problems of deflation. It shifts the deflator from the spending side, including imports, (*YD* is deflated by the deflator for personal consumption expenditures (*PCE*) in *NIPA*) to the product side, including exports. However, this can easily be undone by constructing a deflator for domestic absorption (*NNP* minus exports plus imports) and using it rather than the *NNP* deflator on nominal *NNP*. Another element of data arrangement is to add net investment in consumer durables (*CD*) to *PS* because *CD* is generally regarded as a use of personal saving. Making this change consistently would involve an imputation for the net use value of consumer durables to *NIPA NNP*. Although it is not certain that the size of $PS+CD$ is entirely independent of planned near-term vs. long-term uses of saving on account of possible complementarities between durable and nondurable goods in consumption or in view of downpayment requirements and borrowing constraints, this preparation is not controversial. In addition, the statistical discrepancy in *NIPA* and the item, wage accruals less disbursements, are sometimes added to $PS+CD$ and for good reasons (as I have explained elsewhere).

There is also a growing conviction that corporate saving (*CS*) should be added and that net private saving should be analyzed in toto because of the almost perfect substitutability between the personal and corporate saving components first noted by Martin Feldstein. Recent refinements of that view are that the degree of substitution between *CS* and $PS + CD$ may be less than 1:1 to the extent retained earnings increase the value of the firm only in the ratio of the marginal q to 1. While the marginal q is unobservable, it would have declined by around 50 percent over the past thirteen years if it had retained a fixed relation to the average q of nonfinancial corporations. Because the "offset" coefficient on *CS* may otherwise be significantly less than one and variable, it may be advisable to include qCS as one of the explanatory variables of $PS+CD$. Estimating the latter and *CS* as a simultaneous system may be preferable to estimating a single equation for their sum, total private saving.

Using a multiple-equation framework has the additional advantage of reducing or eliminating simultaneity bias and of allowing one to focus on factors particular to components of private saving such as *CS*. For instance, not only taxes but also dividends appear to be based on reported profits before tax that are generally still not adjusted by businesses for accounting distortions that are due to inflation. As a result, a surge of inflation that raises the

(negative) IVA reduces $NIPA$ CS by an amount that is equal to about two-thirds of the IVA. A decline in the rate of growth of potential output, such as the one percentage point fall that is widely estimated to have occurred in the annual rate since the mid-1960's, also has an effect on CS. Implementing the reduced rate of net investment associated with lower growth requires less internal financing in relation to NNP if the financial structure of corporations is to remain unchanged. It is useful to analyze how factors such as inflation and potential output growth may enter into the sectoral distribution of saving even if changes in CS are in large part offset by changes in $PS+CD$. The uses of private saving for different kinds of investment—residential, business, foreign, and CD—may not be entirely independent of its sources. Assuming, however, that the coefficient found on qCS would be close to -1 in the $(PS+CD)$ equation in system estimation, $PS+CD+qCS$ may be the appropriate concept of private saving to analyze in a single equation framework for some purposes. That framework will also be used to illustrate recent disputes about the role of fiscal variables in private saving equations.

A few years ago Paul David and John Scadding advanced the hypothesis that private saving may be independent of taxes minus transfers $(TAX-TRANS)$ because the private sector considers tax-financed government expenditures as belonging to the category of consumption. Because they regard government consumption and private consumption as interchangeable, "ultra-rational" households then substitute net taxes for PCE at a one-to-one rate according to this argument. Since disposable income YD equals $NNP-CS-(TAX-TRANS)$ (after adding the statistical discrepancy and netting interest paid by consumers to businesses against the corresponding receipt to adjust $NIPA$ YD), David and Scadding would thus expect the coefficient on $(TAX-TRANS)$ to be zero in the private saving and minus one in the personal consumption equation. Although they dealt with gross private saving in relation to GNP, the analogous basic form of the net saving relation implied by them is $PS+CD+CS=a_0NNP$, where a_0 is the marginal and average propensity to save out of NNP. Since $PCE\equiv YD-PS$, the corresponding consumption relation for nondurables reduces to

$$(1)\quad PCE-CD=(1-a_0)NNP-(TAX-TRANS)$$

An alternative possibility, one that is close to the thinking of Robert Barro and others who hold to the (non-)Ricardian, allegedly rational equivalence of government debt and taxes, is to claim that even private saving is too narrow and that national saving should be used instead. This amounts to adding government saving GS, which is the $NIPA$ surplus or deficit $(-)$ of all levels of government combined, to private saving so that $PS+CD+CS+GS=a_1NNP$. Note that $TRANS$ is defined to include all government expenditures other than government purchases of goods and services (GP) though it, of course, also excludes federal grants-in-aid to state and local governments to prevent double counting in the consolidation. Then $GS\equiv TAX-TRANS-GP$. Recalling that $PCE\equiv YD-PS$ and substituting for YD and PS yields the alternative consumption relation

$$(2)\quad PCE-CD=(1-a_1)NNP-GP$$

This formulation captures the essence of the David-Scadding hypothesis, that government consumption is substituted for PCE, rather more forthrightly than identifying government consumption with net taxes. In fact, $TAX-TRANS$ and GP differ not only cyclically but also on the average over cycles as deficits have exceeded surpluses by a substantial margin.

One may next allow that GS and CS may both be substituted for $PS+CD$ at a rate less than 1:1. This concession has been justified by Barro on the ground that unit changes in CS and GS reflect lesser changes in permanent income (because of their transitory nature or the inefficiency with which they are used?). The resulting equation, $PS+CD+\alpha CS+\beta GS=a_2NNP$, with

$(\alpha, \beta < 1)$, yields

$$(3) \quad PCE - CD = (1 - a_2)NNP - \beta GP - (1 - \beta)(TAX - TRANS) - (1 - \alpha)CS$$

If NNP is now decomposed into its trend or permanent income value and cyclical effects represented by the potential output gap or the unemployment rate, one has a fairly general, though highly simplified, form of the aggregate consumption function for nondurables.

That form has been constrained by different researchers in particular ways. Assuming that $\alpha = \beta = a_2$ and that GP does not belong in the last $(PCE - CD)$ equation yields the old disposable income formulation in which a_2 is the marginal and average propensity to save out of YD. Requiring $\alpha = 1$ and $\beta = 0$ yields the basic form implied by David and Scadding, with a_2 equal to a_0. Finally, assuming that corporations act as efficient accumulation agents for households and that GP belongs but $(TAX - TRANS)$ does not because of the equivalence of government debt and taxes amounts to the assertion that $\alpha = \beta = 1$, so that a_2 equals a_1.

Feldstein has pointed out that estimating any of these single equations to determine which of the prior coefficient constraints is more appropriate is likely to involve serious biases because not just NNP but also CS and GS are cyclical. The former leads and the latter lags the reference cycle only slightly. In fact, the cyclicality of GS, that arises from TAX rising and falling with business activity and $TRANS$ moving against it, is extremely pronounced and predictable. Over cycle averages, however, TAX and $TRANS$ have grown by almost equal amounts in relation to NNP. Arguing along permanent income lines, private saving rather than PCE should thus bear the brunt of whatever temporary variations have occurred. However, since the ratio of taxes to NNP is barely cyclical while that of transfers to NNP is strongly counter-cyclical, $TRANS$ accounts for most of the effects of changes in $TAX - TRANS$ over the cycle. This may explain why a very high propensity to save out of transfer payments has been found in a number of econometric studies, the results of which can easily be misinterpreted.

The expansion of the welfare state through equal increases in taxes and transfer payments has been a basic feature of fiscal evolution since the Korean War. Higher taxes on some and increased transfers to others may lead the former group to associate increased future benefits with the higher taxes they pay now. These additional future benefits, though more conditional and constrained, are quite similar to those that could be afforded by private saving. The implication is that, contrary to David and Scadding, increased taxes may be substituted for personal saving providing for future consumption rather than for current consumption. Constraining the coefficients on TAX and $-TRANS$ to be equal precludes one from accounting for this without the aid of additional variables such as net social insurance wealth. Divided by disposable income, the latter rises monotonically from around 1.0 in 1954 to 2.4 in 1977 on a cyclically adjusted basis according to recent estimates. By contrast, the ratio of private net worth to disposable income is trendless. Given that other variables which rise or fall progressively over time, such as the expected real after-tax rate of return on private net worth or one of its components, may be included also, the coefficient estimated on net social insurance wealth may be extremely sensitive to the remainder of the specification in time series estimation. However, the logic supporting a negative effect on private saving, *ceteris paribus*, is persuasive. The countercyclically determined behavior of $TRANS$ would make it inappropriate to cite a positive coefficient on $TRANS$ that exceeds the negative coefficient on TAX as empirical evidence to the contrary.

Not only has the noncyclical pattern of fiscal evolution been stable over recent decades but the customary cyclical responses reflected in GS has displayed a high degree of regularity as well. In spite of the policy

drama that has surrounded each successive tax cut or inflation adjustment act at least since 1964, the net outcome of automatic and discretionary, temporary and permanent fiscal responses has been repetitive over the past twenty-four years, regardless of the labels attached to particular components of change. Faced with this stability of the political rule that governs the size of the government deficit relative to *NNP* that is customarily regarded as appropriate in view of the economic conditions prevailing, one may wonder whether variations in *GS*, properly scaled, can possibly convey anything other than cyclical effects on private saving in time-series for the post-Korean period. In this as in other respects, resolution of the conflicts between the alternative suggested ways in which the basic income argument in saving functions should be expanded beyond (permanent) disposable income is thus unlikely to come from single equation, barely reduced estimates.

The works of all the authors mentioned in this paper are identified in the author's contribution to a conference volume edited by Henry Aaron and Joseph Pechman. Several pertinent discussions have also appeared in the book edited by me especially in the chapters by Willem Buiter and Tobin, and Bulent Gultekin and Dennis Logue. Detailed treatments of the demographic effects on private saving that may be due to the aging, reduced fertility, and increased labor force participation of the *U.S.* population are found in the contributions of Charles Lieberman and Paul Wachtel and Robert Schmitz to the last volume. While these slow-moving variables may be more nearly exogenous than any of the other explanatory variables included in single equation estimates of private saving, they add to the list of time-linked variables whose influence is difficult to separate in aggregate series. On the other hand, disaggregation by cohorts is still impeded by the paucity of reliable survey data on the market value of the assets and liabilities of households that can be tracked consistently over time. Furthermore, linking data on the change in net worth to *NIPA* saving and hence to the investment side of the economy presents a major unresolved problem, as pointed out at the beginning of this article.

REFERENCES

Henry J. Aaron and Joseph A. Pechman, *The Economic Effects of Taxation* (Title preliminary), Washington forthcoming 1980.

Robert J. Barro, *The Impact of Social Security on Private Saving: Evidence from U.S. Time Series, With a Reply by Martin Feldstein,* Washington 1978.

P. A. David and J. L. Scadding, "Private Savings: Ultrarationality, Aggregation, and 'Denison's Law'," *J. Polit. Econ.*, Mar./Apr. 1974, Part I, *82*, 225–49.

M. S. Feldstein, "Corporate Taxation and Dividend Behavior," *Rev. Econ. Stud.*, Jan. 1970, *37*, 59–72.

George M. von Furstenberg, *Social Security Versus Private Saving,* Cambridge, Mass., 1979.

Tax Rules and the Mismanagement of Monetary Policy

By Martin Feldstein*

It is now widely agreed that the mismanagement of monetary policy over the past fifteen years has been a major cause of our current obstinately high rate of inflation. I believe that an important source of this monetary mismanagement has been the failure of the monetary authorities (and of economists in general) to understand how the interaction between inflation and our tax rules influences the effects of monetary policy.

More specifically, as I shall explain in the present paper, I believe that the monetary authorities' failure to recognize the implications of the fiscal structure has caused them to underestimate just how expansionary monetary policy has been. Moreover, because of our fiscal structure, attempts to encourage investment by an easy-money policy have actually had an adverse impact on investment in plant and equipment. The conventional prescription of "easy money and a tight fiscal position" has been an unfortunate guide for macro-economic policy. The switch to floating exchange rates and the relaxation of some of the old restrictions on financial institutions have made it even more important to reject this conventional prescription and to pursue instead a policy mix of "tight money and positive fiscal incentives."

I. Misjudging Monetary Tightness

During the dozen years after the 1951 accord between the Treasury and the Fed, the interest rate on Baa bonds varied only in the narrow range between 3.5 and 5 percent. In contrast, the past fifteen years have seen the Baa rate rise from less than 5 percent in 1964 to more than 10 percent at the beginning of 1979. It is perhaps not surprising therefore that the monetary authorities, other government officials, and many private economists have worried throughout this period that interest rates might be getting "too high." Critics of what was perceived as "tight money" argued that such high interest rates would reduce investment and therefore depress aggregate demand.

Against all this it could be argued, and was argued, that the *real* interest rate had obviously gone up much less. The correct measure of the real interest rate is of course the difference between the nominal interest rate and the rate of inflation that is *expected* over the life of the bond. A common rule of thumb approximates the expected future inflation by the average inflation rate experienced during the preceding three years. In 1964, when the Baa rate was 4.8 percent, this three-year rise in the *GNP* deflator averaged 1.4 percent; the implied real interest rate was thus 3.4 percent. By the beginning of 1969, when the Baa rate was 10.0 percent, the rise in the *GNP* deflator for the previous three years had increased to 6.2 percent, implying a real interest rate of 3.8 percent. Judged in this way, the cost of credit has increased only slightly over the fifteen-year period.

All of this ignores the roles of taxes. Since interest expenses can be deducted by individuals and businesses in calculating taxable income, the net-of-tax interest cost is very much less than the interest rate itself. Indeed, since the *nominal* interest expense can be deducted, the *real* net-of-tax interest cost has actually varied inversely with the nominal rate of inflation. *What appears to have been a rising interest rate over the past fifteen years was actually a sharply falling real after-tax cost of funds.* The failure to recognize the role of taxes prevented the mone-

*Harvard University and the National Bureau of Economic Research. The views expressed here are my own and should not be attributed to any organization.

tary authorities from seeing how expansionary monetary policy had become.

The implication of tax deductibility is seen most easily in the case of owner-occupied housing. A married couple with a \$30,000 taxable income now has a marginal federal income tax rate of 37 percent. The 10 percent mortgage rate in effect at the beginning of 1979 implied a net-of-tax cost of funds of 6.3 percent. Subtracting the 6.2 percent estimate of the rate of inflation leaves a real net-of-tax cost of funds of only 0.1 percent. By comparison, the 4.8 percent interest rate for 1964 translates into a 3.0 percent net-of-tax rate and a 1.6 percent real net-of-tax cost of funds. Thus, although the nominal interest rate had more than doubled and the real interest rate had also increased, the relevant net-of-tax real cost of funds had actually fallen from 1.6 percent to only 0.1 percent.

As this example shows, taking the effects of taxation into account is particularly important because the tax rules are so nonneutral when there is inflation. If the tax rules were completely indexed, the effect of the tax system on the conduct of monetary policy would be much less significant. But with existing tax rules, the movements of the pre-tax real interest rate and of the after-tax real interest rates are completely different. I think that monetary policy in the last decade was overly expansionary because the monetary authorities and others believed that the cost of funds was rising or steady when in fact it was falling significantly.

The fall in the real after-tax interest rate has caused a rapid increase in the price of houses relative to the general price level (see, for example, Patric Hendershott and and Sheng Cheng Hu) and has sustained a high rate of new residential construction. There were, of course, times when the ceilings on the interest rates that financial institutions could pay caused disintermediation and limited the funds available for housing. To that extent, the high level of nominal interest rates restricted the supply of funds at the same time that the corresponding low real after-tax interest cost increased the demand for funds. More recently, the raising of certain interest rate ceilings and the development of mortgage-backed bonds that can short circuit the disintermediation process have made the supply restrictions much less important and have therefore made any interest level more expansionary than it otherwise would have been.

The low real after-tax rate of interest has also encouraged the growth of consumer credit and the purchase of consumer durables. More generally, even households that do not itemize their tax deductions are affected by the low real after-tax return that is available on savings. Because individuals pay tax on nominal interest income, the real after-tax rate of return on saving has become negative. It seems very likely that this substantial fall in the real return on savings has contributed to the fall in the personal saving rate and the rise in consumer demand.[1]

For corporate borrowers, the analysis is more complex because inflation changes the effective tax rate on investments as well as the real net-of-tax interest rate. More specifically, historic cost depreciation and inventory accounting rules reduce substantially the real after-tax return on corporate investments (see my forthcoming paper with Laurence Summers). An easy-money policy raises the demand for corporate capital only if the real net cost of funds falls by more than the return that firms can afford to pay. This balance between the lower real net interest cost and the lower real net return on investment depends on the corporation's debt-equity ratio and on the difference between the real yields that must be paid on debt and on equity funds. It is difficult to say just what has happened on balance. In a preliminary study (1978), Summers and I concluded that the rise in the nominal interest rate caused by inflation was slightly less than the rise in the maximum interest rate that firms could afford to pay. However, this analysis made no allowance for the

[1]Although the response of household saving to even a compensated change in the interest rate is theoretically ambiguous (see my 1978 paper), plausible parameter values and some econometric evidence support a positive saving elasticity (see Michael Boskin; Summers).

effect of inventory taxation or for the more complex effects of inflation on equity yields that I have recently investigated in my 1979a paper. My current view is that on balance monetary policy reduced the demand for business investment at the same time that it increased the demand for residential investment and for consumption goods.

It is useful to contrast the conclusion of this section with the conventional Keynesian wisdom. According to the traditional view, monetary expansion lowers interest rates which reduces the cost of funds to investors and therefore encourages the accumulation of plant and equipment. This statement is wrong in three ways. First, a sustained monetary expansion raises nominal interest rates. Second, although the interest rate is higher, the real net-of-tax cost of funds is lower. And, third, the lower cost of funds produced in this way encourages investment in housing and consumer durables (as well as greater consumption in general) rather than more investment in plant and equipment.

II. The Correct Mix of Monetary and Fiscal Policies

There is widespread agreement on two central goals for macro-economic policy: 1) achieving a level of aggregate demand that avoids both unemployment and inflation; and 2) increasing the share of national income that is devoted to business investment. Monetary and fiscal policy provide two instruments with which to achieve these two goals. The traditional view of the effect of monetary policy has led to the conventional prescription of easy money (to encourage investment) and a tight fiscal policy (to limit demand).

This policy mix could in principle achieve its two goals. A government surplus would permit a reduction in the supply of government liabilities (money and bonds) and would thereby facilitate increased capital accumulation. The required change in the interest rate would depend on the relative interest sensitivities of the market demand for bonds and money. In the likely case in which the demand for money is relatively inelastic, the government surplus must be accompanied by a lower rate of interest and the substitution of real capital for government bonds.[2]

Unfortunately, the traditional prescription of easy money and a tight fiscal position is almost bound to fail in practice because of the political difficulty of achieving and maintaining a government surplus. During the past twenty years, there have not been any two successive years in which the federal government budget showed a surplus. As a result, the pursuit of an easy-money policy has produced inflation. Although the inflationary increase in the money supply did reduce the real after-tax cost of funds, this only diverted the flow of capital away from investment in plant and equipment and into owner-occupied housing and consumer durables. By reducing the real net return to savers, the easy-money policy has probably also reduced the total amount of new saving.

The inappropriateness of the traditional policy mix reflects not only its overoptimistic view about the feasibility of government surpluses but also its overly narrow conception of the role of fiscal policy. In the current macro-economic tradition, fiscal policy has been almost synonymous with variations in the net government surplus or deficit and has generally ignored the potentially powerful incentive effects of taxes that influence marginal prices.

A more appropriate policy mix for achieving the dual goals of balanced demand and increased business investment would combine a tight-money policy and fiscal incentives for investment and saving. A tight-money policy would prevent inflation and would raise the real net rate of interest. Although the higher real rate of interest would tend to deter all forms of investment, specific incentives for investment in plant and equipment could more than offset the higher cost of funds. The combination of the higher real net interest rate and the targeted investment in-

[2]See my forthcoming (1980) paper for a formal model of the relation of fiscal policy, monetary policy, and capital formation in a fully employed economy.

centives would restrict housing construction and the purchase of consumer durables while increasing the flow of capital into new plant and equipment. Since housing and consumer durables now account for substantially more than half of the private capital stock, such a restructuring of the investment mix could have a substantial favorable effect on the stock of plant and equipment.

A rise in the overall saving rate would permit a greater increase in business investment. The higher real net rate of interest would in itself tend to induce such a higher rate of saving. This could be supplemented by explicit fiscal policies that reduced the tax rate on interest income and other income from saving.

Switching from an easy-money policy to a policy mix with high real interest rates would have a further advantage. Because of the current system of floating exchange rates, a rise in the real interest rate would cause an appreciation of the dollar which would in turn reduce the price level directly (see, for example, Rudiger Dornbusch and Paul Krugman). With less than perfect international capital mobility, higher interest rates could persist and would tend to attract some inflow of foreign capital that would further augment investment in the United States.

III. Macro-Economic Importance of the Fiscal Structure

The misjudgement of monetary tightness and the advocacy of an inappropriate policy mix suggest the importance of recognizing that *the fiscal structure of our economy is a key determinant of the macro-economic equilibrium and therefore of the effect of monetary policy*. Conventional macro-economic analyses that ignore the fiscal structure (or that assume that all taxes are equivalent to lump sum taxes) can be seriously misleading.[3] The fact that the real interest rate can simultaneously rise on a pre-tax basis and fall on a net-of-tax basis shows that fiscal effects with the existing *U.S.* tax law are qualitatively as well as quantitatively important.

The common tendency to ignore the tax structure in macro-economic analysis is due at least in part to the fact that taxes were much less important at the time that the current models of macro-economic analysis were developed. When John Maynard Keynes' *General Theory* was first published, less than 5 percent of American families were affected by the income tax and the median tax rate among those who paid tax was less than 5 percent. The greater current significance of the fiscal structure reflects not only the growth of the income tax, but also the increased importance of social insurance programs like unemployment insurance and Social Security.

The tax structure is particularly important as a cause of the macro-economic nonneutrality of inflation. Irving Fisher's (1930) famous conclusion that inflation raises the nominal interest rate but leaves the real rate unchanged is appropriate for an economy with no taxes but not for an economy in which nominal interest payments are reflected in income and profits that are subject to substantial marginal tax rates. I have shown elsewhere (1976) that in an economy with economic depreciation and a 50 percent tax rate, each 1 percent increase in the inflation rate raises the nominal interest rate by 2 percent. With other depreciation rules, the effect of the inflation rate on the interest rate can be more than one-to-one or less than one-to-one (see my paper with Jerry Green and Eytan Sheshinski).

More generally, the tax structure determines how inflation influences the real values of corporate equities (see my 1979a and forthcoming papers and Hendershott's paper), of residential real estate (see Hendershott and Hu), and of such "store of value" assets as land and gold (see my 1979b paper). As David Hartman has shown, the tax rules may also induce international capital flows in response to changes in inflation that would have no real effect in the absence of taxation.

[3]Ignoring the tax structure is analogous to ignoring the international aspects of domestic economic equilibrium; there are issues for which both taxes and international aspects can be ignored but there are others for which doing so would be very inappropriate.

IV. Conclusions

This paper has emphasized the importance of the fiscal structure as a determinant of the macro-economic equilibrium. It discussed the desirability of substituting a policy of tight money and positive fiscal incentives for the traditional goals of easy money and fiscal restraint. More specifically, it identified the failure to recognize the effects of the tax rules as an important reason for the mismanagement of monetary policy. Although the guidance of short-run monetary policy may now give more weight to controlling the money supply and credit aggregates, interest rates will almost certainly continue to influence the determination of longer-term monetary policy. The correct interpretation of the relation between interest rates and inflation therefore remains important and will become even more important if the authorities attempt to shift to a policy of tighter money and an altered tax structure.

REFERENCES

M. J. Boskin, "Taxation, Saving, and the Rate of Interest," *J. Polit. Econ.*, Apr. 1978, *86*, S3–S27.

R. Dornbusch and P. Krugman, "Flexible Exchange Rates in the Short Run," *Brookings Papers*, Washington 1976, *3*, 537–84.

M. Feldstein, "Inflation, Income Taxes and the Rate of Interest: Theoretical Analysis," *Amer. Econ. Rev.*, Dec. 1966, *56*, 809–20.

______, "The Rate of Return, Taxation, and Personal Saving," *Econ. J.*, Sept. 1978, *88*, 482–87.

______, "Inflation, Tax Rules, and the Stock Market," Nat. Bur. Econ. Res. work. paper no. 403, Cambridge, Mass. 1979a.

______, "The Effect of Inflation on the Prices of Land and Gold," Nat. Bur. Econ. Res. work. paper no. 296, Cambridge, Mass. 1979b.

______, "Fiscal Policies, Inflation and Capital Formation," *Amer. Econ. Rev.*, forthcoming 1980.

______, "Inflation and the Stock Market," *Amer. Econ. Rev.*, forthcoming.

______, **J. Green, and E. Sheshinski,** "Inflation and Taxes in a Growing Economy with Debt and Equity Finance," *J. Polit. Econ.*, Apr. 1978, *86*, S53–S70.

______ **and L. Summers,** "Inflation, Tax Rules and the Long-Term Interest Rate," *Brookings Papers*, Washington 1978, *1*, 61–100.

______ **and** ______, "Inflation and the Taxation of Capital Income in the Corporate Sector," *Nat. Tax J.*, Dec. 1979, *32*, 445–70.

D. Hartman, "Taxation and the Effects of Inflation on the Real Capital Stock in an Open Economy," *Int. Econ. Rev.*, June 1979, *20*, 417–25.

P. Hendershott, "The Decline in Aggregate Share Values: Inflation and Taxation of the Returns from Equities and Owner-Occupied Housing," Nat. Bur. Econ. Res. work. paper no. 370, Cambridge, Mass. 1979.

______ **and S. C. Hu,** "Inflation and the Benefits from Owner-Occupied Housing," Nat. Bur. Econ. Res. work. paper no. 383, Cambridge, Mass. 1979.

L. H. Summers, "Tax Policy in a Life Cycle Model," Nat. Bur. Econ. Res. work. paper no. 302, Cambridge, Mass. 1978.

DISCUSSION

MARVIN KOSTERS, American Enterprise Institute for Public Policy Research: An enormous amount of research has been done on factors influencing labor supply, at least for the conventional measures of labor supply—hours of work and labor force participation. One of the most notable advances in this area of research has been the introduction of nonlinearity into models and estimating techniques in order to analyze the effects of taxes. It is interesting, however, that, at least in a qualitative sense, the results of early research suggesting "small" substitution effects for adult males and considerable sensitivity for married women have not been altered greatly by new and sophisticated analyses of conventional data and of information from income maintenance experiments.

It is heartening that labor supply incentives have received so much attention. What is less encouraging, however, is that there remains considerable uncertainty about how much confidence can be placed on the evidence that has been developed. One source of uncertainty has conceptual roots. Income from nonemployment sources has usually been treated as largely exogenous in labor supply estimation. But in a life cycle model of asset accumulation, hours and assets can be expected to be simultaneously determined by similar economic forces; that is, assets, and hence nonemployment income, are endogenous. As it happens, the problems this creates for estimation are likely to be more critical for adult males than for married women. Consequently, in my view, we can place more confidence in the estimates that have been developed for married women (where the evidence indicates economically significant responses) than for adult males (where the evidence suggest only small effects).

In his paper, Harvey Rosen discusses some questions on which there is little research evidence, such as the adequacy of conventional measures of labor supply. In his discussion of "intensity of work," he suggests that "effective" hours of work may be more responsive to wages than conventionally measured hours. If this is so, since much of the current evidence is based on conventional hours measures, effective hours and labor supply may be more sensitive to incentives than is suggested by most of the evidence derived from analysis of hours of work.

The choice of retirement age is one element in the broader class of questions that arise in connection with such policies as the earnings test for Social Security recipients and the effects of programs such as unemployment insurance, disability insurance, trade adjustment assistance, and the treatment of benefits from these various programs under the income tax law. These various programs have potentially important effects on labor supply, and analyses of them can potentially provide additional insight into labor supply sensitivity.

A third dimension of labor supply discussed in the paper is labor quality in the context of human capital investment. For reliable evidence on this question I think a fairly comprehensive approach is necessary. Education and training are strategic decisions, with students and their parents participating jointly in the decisions. They are strategic in the sense that they involve actually developing attributes that are usually taken as given in labor supply studies. What needs to be taken into account in analysis of such lifetime choices is not only differences in before-tax wages (which are themselves likely to be affected by the tax system), but also the incidence of taxes for different earnings profiles. To address the entire distribution of human capital investment and earnings, account would have to be taken of the terms under which income support and services are available at the low end of the distribution. Essentially, what is at issue is differences in real after-tax and after-transfer incomes that are associated with differences in human capital investment.

Finally, one aspect of the labor supply evidence that is not discussed in Rosen's paper—but which I believe may be of greater importance than some that are discussed—is the role of customary work prac-

tices, the legal framework establishing "normal" working hours, and general attitudes toward working hours that are associated with them. These are essentially stable and given in developing all of the empirical evidence. It is possible—perhaps likely—that the evidence would show more responsiveness if these customs and laws were permitted to vary instead of being fixed for the context where the data were generated. This is important because if the incentives built into our tax and transfer policies lead to broad changes in desired labor supply, customary practices will shift over time, and legislative change is likely to confirm or even to speed up such changes. Attitudes and customs may have important effects on normal hours of work patterns, on human capital investment decisions, and on tendencies to rely more extensively on transfers instead of earnings. We have little evidence at present on how important these factors may be. The possible constraining effect of conventional hours of work is difficult to explore empirically, but I think the influence of such conventions is a major source of uncertainty about how much confidence can be placed on the current evidence from research on labor supply sensitivity.

JOHN B. SHOVEN, Stanford University and National Bureau of Economic Research: It is difficult to quarrel with what I view as the three main points of George von Furstenberg's paper: First, it is clear that for periods as short as one year, asset revaluations loom very large relative to net saving, and therefore such revaluations are the dominant factor determining the change in net wealth. This certainly adds a great deal of noise to a model of saving which is based on the difference between desired and actual wealth.

Second, it certainly is difficult to construct an acceptable proxy for the expected after-tax real rate of return on the total household portfolio. Overall, I am less pessimistic than von Furstenberg about the possibilities of doing so, however. In the long run, certainty equivalent after-tax yields tend to equalize, and I believe that this equalized rate of return can be measured subject to reasonable standard errors.

Third, I also agree that a proper treatment of saving should be a comprehensive one. As Martin Feldstein and others have previously argued, individuals may pierce the corporate and government veils in making their private saving decisions. A proper treatment of savings thus should take into account corporate and government saving as well as consumer durables and financial accumulation.

I disagree with the author's argument that only (Tobin's) "q" times corporate savings affect private savings because of the fact that financial markets only fractionally capitalize retained earnings. What is needed for this argument is marginal q, the increase in the financial valuation of the firm when it retains an extra dollar. But von Furstenberg states that he is forced to use average q for empirical reasons. The link between the average valuation of the firm and the marginal valuation of retained earnings is very precarious. For example, when the price of energy goes up, the average q is likely to fall (the value of existing assets will decline), while marginal q may rise. Financial markets will reflect the incentive to replace existing, relatively energy-intensive facilities with new, more energy-conserving capital stock. An additional argument against the treatment in this paper is that models in which dividends can be costlessly adjusted and firms can repurchase their own equity typically have the property that a firm acting in the interests of its stockholders will act so as to force marginal q to unity. This entire issue regarding the inclusion of retained earnings in corporate savings could have been avoided, had the author defined corporate savings to be the real appreciation of corporate financial securities net of new issues.

Finally, I would warn the reader of this paper that it offers useful commandments for research on private saving, but following them is not easy. Even the author, in his contribution to *The Economic Effects of Taxation* (Henry Aaron and Joseph Pechman, editors) does not completely

adhere to a comprehensive view of saving, where each component depends on saving in other forms or elsewhere in the economy. In that paper he sets up a recursive structure in which (loosely) government savings and taxes are first set, they in turn affect corporate saving, which (together with the government accounts) affect private saving. Clearly, one can argue whether this recursive structure is correct (for example, whether the government "goes first"); it would seem that a more complete interactive system would be superior.

In conclusion, I cannot fault the principles stated in this paper. However, it does not give much guidance on how they are to be applied, and that task should not be underestimated.

ALAN S. BLINDER, Princeton University: Martin Feldstein argues that monetary policy has been too permissive during the last ten to fifteen years and that a better policy would have been tighter money coupled with strong pro-investment tax incentives. I find his argument unconvincing, for several reasons.

First, Feldstein claims that monetary policy in the last decade was "overly expansionary" because the Fed failed to realize that the *rising* nominal interest rates represented *falling* after-tax real interest rates, and thus "loose money."

I do not see such a clear and consistent pattern. The period began with a "credit crunch" in 1969, followed by appropriately easier money during the 1970–71 recovery. Next came the well-documented pre-election stimulation of the economy, where I happily join Feldstein (and just about everyone else) in condemning the Fed for excessive money creation. Thereafter, however, I find it hard to imagine which were the "overly expansionary" periods (ignoring a few "blips" in money growth rates). During the recession year of 1974, for example, the money supply increased 5.3 percent (while the *GNP* deflator rose 9.7 percent). I cannot recall anyone characterizing monetary policy during the 1975–76 recovery period as excessively expansionary, and money growth in 1978 was merely 4.3 percent. This leaves only 1977 as possibly overly expansionary. I do not read this record as one of profligacy, and find Robert Hall's characterization of postwar policy more accurate: "Anti-recessionary policy has been timid and even perverse in the postwar decades."

According to Feldstein, the low real interest rates produced by inflation biased investment spending in favor of housing and away from business fixed investment. While this is often claimed, it has not been proven to my satisfaction. A bias in favor of homebuilding certainly *is* written into our tax structure; and it *does* seem true that inflation raises the rate of return on housing. But we should remember that inflation also exacerbates certain capital market and liquidity constraints that deter homebuilding. First, the disintermediation process that Feldstein dismisses rather cavalierly operated with a vengeance through most of the fifteen-year period that he considers. High *nominal* interest rates did indeed deter investment in housing. Second, the institutional convention that home mortgages be written with level payments in *nominal* terms means that the real payback on a mortgage is very heavily "front loaded" when inflation is high. Such heavy initial financial burdens no doubt dissuade many potential homebuyers.

But what of the alleged bias *against* business fixed investment? As Feldstein notes, our unindexed tax system means that inflation erodes the real value of depreciation allowances, and thus reduces the rate of return on corporate investment, but also lowers real after-tax borrowing costs. Theory does not suggest that inflation damages capital formation. Does the data suggest this? During the nine years from 1956 through 1964, when the inflation rate (*GNP* deflator) averaged only 2 percent, business fixed investment averaged 9 percent of *GNP*. During the eight years from 1965 through 1972, when the average inflation rate accelerated to 4.1 percent, the ratio of business fixed investment to *GNP rose* to 10.3 percent. During the high inflation years 1973–78, business fixed investment averaged 10 percent of *GNP*. Over the en-

tire 1956–78 period, the simple correlation between inflation and the ratio of business fixed investment to *GNP* is *positive* .53, not negative as Feldstein suggests.

Next, Feldstein asserts that low real after-tax interest rates have been an important cause of low personal savings. Again the case must be empirical, not theoretical. But Michael Boskin's study, which Feldstein cites, has been disputed by Philip Howrey and Saul Hymans, George von Furstenberg, and others; and other investigators have obtained negligible interest elasticities of saving. No one at this time can confidently claim that a strong saving response to higher interest rates is well established.

Finally, Feldstein observes that one way to raise the share of business investment in *GNP* is to keep the government budget in surplus and pursue an easy monetary policy; but he claims that chronic budget deficits have precluded this strategy. This claim seems based on faulty accounting. Just as inflation produces "illusory" profits (that we then tax), so also does inflation produce illusory budget deficits. Specifically, the federal government each year reaps a large capital gain on its outstanding debt, which is not reckoned into the budget. Put differently, the budget counts *nominal* interest expense, when it should count *real* interest expense.

In addition to this, it is clear that if we are worried about government debt "crowding" corporate borrowers out of the bond market, the appropriate index in a *growing* economy is the ratio of public debt to *GNP* (or to total wealth)—a ratio that has *fallen* steadily since World War II. Rough corrections for these two factors lead to the conclusion that the federal budget has been more often in the black than in the red.

It is clear by now that I do not buy Feldstein's argument. Why, then, has business fixed investment been weak? I suggest a rather different explanation, which Robert Hall expressed for me: "The principal source of foregone capital formation has been our failure to do anything about recessions, not our active use of anti-investment stimulative policies." To state the problem in this way is to suggest a remedy: use fiscal and monetary policy to combat recessions. I hope Feldstein will join me in recommending such policies for the 1980's, for I am sure they would lead to much stronger capital formation.

CONSEQUENCES OF THE GROWTH OF THE TWO-EARNER FAMILY

Urban Land Use and the Growth in Two-Earner Households

By JANICE FANNING MADDEN*

Two-earner households differ from one-earner, husband-wife households in other ways than the number of earners. They have higher money incomes; they are smaller and also less likely to include young children; they have less adult "leisure" time. Urban residential location choices have traditionally been explained in terms of tradeoffs between housing consumption and job accessibility. These differences in income, household composition, number of workplaces, and available leisure time thus suggest that the residential and work location choices, and housing demands of two-earner households differ from those of other households.

Because two-earner households have higher money incomes and because recent empirical studies have found the income elasticity of housing demand in the United States to be greater than one (see Edwin Mills), two-earner households are expected to consume more housing. Because the housing expenditure savings from a more suburban location increase as housing consumption increases, households who consume larger amounts of housing (i.e., two-earner) are expected to find suburban locations more attractive, *ceteris paribus*.

On the other hand, two-earner households have members commuting to two jobs. This suggests that commuting costs are "doubly" incurred and that access to jobs plays a larger role in their residential location decision. If employment of both spouses tends to be centrally located, the total commuting costs associated with more suburban locations are higher for two-earner than for one-earner households. Alternately, if one spouse is employed at a suburban location (see Michelle White), some suburban locations may offer proportionately lower commuting costs for two-earner than for one-earner households. To the extent that: 1) employment location is determined by residential location and commuting preferences of employees; and 2) either two-earner households locate differently than one-earner households, or employed wives have different commuting preferences than other employees, the increasing labor force participation of married women will also affect the rate of suburbanization of employment.

Finally, besides having higher money incomes and more job locations to access, two-earner households may differ systematically from one-earner households in their preferences for particular housing characteristics such as neighborhood and density.

This paper analyzes the net effect of income, commuting costs, and housing composition on the household's choices of house size, house location relative to employment location, and other housing characteristics for two-earner and one-earner households. In the first section a household decision-making model of location and housing consumption is developed; the second section discusses the data used to estimate the model; the third section discusses the estimation procedure and presents the empirical results; the final section presents the conclusions.

I. The Model

An analysis of the effect of an additional earner on household location choices and

*Associate professor of regional science, University of Pennsylvania. This research was supported by National Institute of Mental Health Grant No. 1-R01-MH-3400-01.

housing demand at a location requires a model which permits more than one job location (i.e., does not assume all jobs are in the center).

For a household, utility U is a function of the amount of housing H and other goods X consumed:

$$(1) \qquad U=f_1(H,X)$$

There is a spatially continuous set of job alternatives available which offer varying wages W as a function of factors Z representing general supply and demand conditions in the labor market, as well as distance of the job from the center t:

$$(2) \qquad W=w(t,Z)$$

Assuming that employers' nonwage costs increase (or revenues decrease) with distance from the center, t would be inversely related to W.

Similarly, per unit housing prices R are dependent on a number of factors Q, representing general supply and demand conditions in the housing market, as well as the distance u of the residence from the city center:

$$(3) \qquad R=r(u,Q)$$

Both theoretical and empirical work indicate that the rental value of housing space is inversely related to u.

Workplace-residence separation is equal to the absolute value of the difference between residential location and job location: $|u-t|$. Time spent commuting is equal to the physical separation divided by the speed of travel S.

The household is thus constrained by the following budget:

$$(4) \qquad X+r(u,Q)H+w(t,Z)\frac{1}{S}|u-t|=WM+V$$

where the price of other goods X is the numeraire set equal to 1, M is work hours on the job, and V is the household's unearned income. Selecting H,X,u,t so as to maximize (1) subject to (4) yields the equilibrium workplace-residence separation (noted by the superscript e)

$$(5) \qquad |u-t|^e=f_2[w(t,Z),r(u,Q),H,M,V]$$

When the household includes an additional earner, an additional work location t_2 and wage rate W_2 can be included in equation (4):

$$(4') \qquad X+r(u,Q)H+w(t_1,Z_1)\frac{1}{S}|u-t|= W_1M_1+V+W_2M_2-w(t_2,Z_2)\frac{1}{S}|u-t_2|$$

The equilibrium workplace-residence separation is selected with respect to H, X, u, t_1, and t_2, and can be defined with respect to either earner's work location:

$$(5') \qquad |u-t_i|^e=f_2[w(t_1,Z_1),w(t_2,Z_2), r(r,Q),H,M_1,M_2,V]$$

At the equilibrium residential location, the household consumes an equilibrium housing consumption bundle H, which is a function of the physical size of the dwelling SIZ and of the quality of the dwelling $QUAL$ as indicated by amenities such as neighborhood, design, public school quality, etc.:

$$(6) \qquad H^e=(SIZ^e,QUAL^e)$$

The household's choices of SIZ and $QUAL$ at their equilibrium location depend on their income and on their tastes as differentiated by their demographic characteristics D:

$$(7) \qquad SIZ^e=f_3(W_1M_1+W_2M_2+V,|u-t|^e,D)$$

$$(8) \qquad QUAL^e=f_4(W_1M_1+W_2M_2+V,|u-t|^e,D)$$

Equations (5), (7), and (8) determine the household's equilibrium housing consumption. (Of course, the household actually selects location and housing bundle com-

TABLE 1—MEAN LABOR MARKET, HOUSING AND FAMILY CHARACTERISTICS FOR HOUSEHOLDS WITH EMPLOYED HEADS, LOCATED WITHIN 30 MILES OF A CITY OVER 100,000: PANEL STUDY OF INCOME DYNAMICS, 1976[a]

	Two-Earner with Children	Traditional[b] with Children	Female Head with Children	Two-Earner Couple	Traditional[b] Couple	Unmarried Women	Unmarried Men
Distance to Job from Residence:							
Head	12.5	11.8	7.7	11.1	10.4	7.5	8.2
	(12.0)	(10.7)	(5.8)	(8.9)	(8.3)	(7.1)	(7.0)
Spouse	7.7			10.1			
	(5.7)			(7.4)			
Housing Characteristics:							
Number of rooms	6.1	5.8	5.0	5.4	5.3	4.1	3.7
	(1.2)	(1.4)	(1.5)	(1.6)	(1.3)	(1.6)	(1.8)
Distance of residence from city center:							
0–5 miles	.18	.30	.35	.29	.25	.41	.42
5–15 miles	.59	.48	.40	.51	.55	.44	.43
over 15 miles	.24	.22	.25	.20	.20	.15	.15
City's population over 500,000	.47	.49	.44	.53	.58	.55	.54
Household Characteristics:							
Number of children	1.9	2.1	1.7				
	(1.1)	(1.1)	(1.0)				
Age of head	37.0	36.8	36.7	38.2	49.9	40.4	30.3
	(9.0)	(11.0)	(9.6)	(14.5)	(13.6)	(15.9)	(11.5)
Family annual income	23240	18429	11953	23920	21739	11489	13395
	(12320)	(8909)	(6199)	(11026)	(11127)	(9794)	(8852)
Moved since 1975	.13	.22	.33	.34	.11	.35	.50
Ever married			.86			.48	.35
Black	.12	.10	.26	.07	.01	.12	.20
Sample Size	165	243	52	162	75	124	94

[a]Numbers in parentheses are standard deviations.
[b]Traditional indicates a household with an employed husband and a nonemployed wife.

position simultaneously. This study focuses on equilibrium outcomes and estimates the tradeoff between size and quality of housing at an equilibrium location because with the available data constraints it is only possible to define quality once the accessibility proportion of housing price is given.)

II. The Data

The data used to analyze housing market and commuting behavior are the 1976 wave of the *Panel Study of Income Dynamics* (*PSID*) collected by the Survey Research Center at the University of Michigan. Although confidentiality restrictions do not permit the data user to know the city in which the family resides, the data are particularly unique as a rich source of information on labor force characteristics, household characteristics, work trip behavior, and housing characteristics.

Table 1 lists the sample sizes, the means, and standard deviations of household characteristics and housing variables for seven categories of households, which reflect the number of adults, number of earners, and presence of children in the household. These households all contain *employed* members. In order to make this a study of urban location decisions, households residing in census sampling units with no cities over 100,000 or who were more than 30 miles from the center of a city with population over 100,000 were eliminated from the sample.

As anticipated, two-earner households have larger total money incomes (although husbands have lower incomes) and occupy larger houses than one-earner, husband-wife households. While two-earner couples tend to reside more centrally than one-earner couples, two-earner families are more suburbanized than one-earner families. However, it is unmarried individuals who reside more centrally than all other household units. Individuals in two-earner households reside farther from their work locations than do individuals of the same sex from other household categories. (It is interesting to

note that although unmarried and married women with children have workplaces at comparable distances from their residences, the disparity in their residential location suggests large differences in their job locations.)

While income and household size have their expected effects on the two-earner household's residential location, Table 1 does not indicate whether income and household size fully account for the difference in one-earner and two-earner households, or whether an additional earner and/or job location in itself differentiates housing outcomes. In order to sort out the effects of an additional earner from those of higher money income and smaller households on the location and housing decisions of two-earner households, it is necessary to analyze empirically the determinants of the household's equilibrium housing consumption.

III. Estimation of the Model

As developed above, the empirical model includes equations (2), (3), (5), (7), and (8) and is estimatable using a two-stage least squares approach. The data do not permit this direct approach, however.

Although the data include the distance of residence from the center city, the population of the city, and the distance of the job from the residence, they do not include the distance or direction of job from the center city. Therefore, equation (2) cannot be directly estimated in the form specified. Rather, this equation is estimated using distance of job from residence $\hat{t}$ for t and also controlling for residential location; that is, equation (2) is estimated separately for each category of city size and distance from center. As $\hat{t}$ reflects distance from a fixed residential location rather than distance from the city center, increases in $\hat{t}$ imply either that employment is more centralized or in a different direction from the city center than residential location is. In either case, a positive relation between wages and work trip length is both expected and found.

Because the data do not contain any measures of housing quality other than price per room and because there is no other information on the housing market, Q is not known and it is not possible to directly estimate equation (3). However, on the basis of abundant empirical and theoretical literature which has found that *ceteris paribus* per unit housing prices are a positive function of city size and a negative function of distance from city center, these variables are entered directly into the estimation of equation (5).

Thus, equation (5) is estimated in the following form:

$$(5'') \quad |u-t|^e = f_1(\hat{W}_{dc}, d, c, C, D, V, \gamma)$$

where $\hat{W}_{dc}$ is the individual's wage as estimated by equation (2) given his or her household's residential location in terms of d, distance from city center, and c, population of city; d and c measure the effect of the rent gradient, given the expected wage at the residential location; C is a vector of labor market characteristics including weekly work hours and months with current employer; D is a vector of household characteristics including marital status, number of children, age, race, and whether residence is owned or has changed in the past year; V is total family income; and γ is the error term. This procedure is tantamount to a two-stage least squares estimations of equations (2) and (5).

Equations (7) and (8) are estimated using $|u-t|^e$ as estimated in equation (5″). Number of rooms in the household's housing unit is used to represent SIZ; the price or value per room of the housing unit is used to represent $QUAL$. Because equation (8) controls for distance from center of city, city size, and distance from jobs with $|u-t|^e$, the variation in price per room reflects the remaining housing characteristics such as lot size, neighborhood quality, and net cost of public goods and services attached to the house.

Finally, the empirical estimation of equations (5″), (7), and (8) is conducted separately for men and women because: 1) sex is correlated with labor force characteristics such as job tenure and location; and 2) sex interacts strongly with the demographic characteristics measured (i.e., while

TABLE 2—DETERMINANTS OF HOUSING CONSUMPTION OF MEN EARNERS

Independent Variables	Dependent Variables: *Ln* of Workplace-Residence Separation	*Ln* of House Size	*Ln* of House Amenities
Relative to Employed Wife:			
Employed wife and children	.075	.121[a]	−.074
Nonemployed wife and children	−.005	.132[a]	−.054
Nonemployed wife	.089	−.029	.077
Estimated wage	1.715[a]		
Residence			
within 5 miles of center	−.097		
15–30 miles from center	.345[a]		
in city of over 500,000	−.214[a]		
Ln of job tenure	−.096[a]		
Ln of weekly hours	−.301[c]		
Black	.641[a]	.006	−.367[a]
Moved in 1975	.154	.036	.165[a]
Age	−.001	.020[a]	.026[b]
Age squared	−.000	−.0002[a]	−.0003[b]
Own home	.135[c]	.128[a]	.170[a]
Ln of household income		.138[a]	.268[a]
Estimated *ln* of worktrip		.033	.109[b]
Wife has			
high school or more		.044	.148[a]
college or more		.939[b]	.164[b]
Intercept	.452	−.402	5.050
MSE	.7561	.0525	.2218

[a]Prob > .99 that coefficient is nonzero.
[b]Prob > .95 that coefficient is nonzero.
[c]Prob > .90 that coefficient is nonzero.

men earners in one-earner households include both married and unmarried men, women earners in one-earner households are almost entirely single.) In order to assess the extent to which two-earner households behave differently than one-earner households, it is necessary to control for both the sex of the earner and the demographic composition of one-earner households. Estimating the model separately for men and for women provides a convenient means for controlling for the sex, labor force, and household demographic interactions.

Table 2 presents the estimated coefficients of the housing consumption model for men earners who are married with wife present. Because previous research has indicated that the marriage market is selective of men with higher earnings potential, Table 2 excludes single men. In order to control for the earnings potential of wives, regardless of their current labor force status, wife's education attainment was used as well as household income in the house size and quality equations.

The independent variables generally behave as expected: increasing wages, lower house prices, and less job tenure significantly increase workplace-residence separation, and increasing household income and age increase both the size and quality of housing consumed.

The household demographic variables are generally statistically insignificant for married men. Differences in the housing consumption of one-earner and two-earner, husband-wife households seem to be fully explained by their differences in money income and fertility. Although men from two-earner couples reside closer to their jobs than men with either nonemployed wives or employed wives and children, and about the same distance from their jobs as men with nonemployed wives and children, none of

TABLE 3—DETERMINANTS OF HOUSING CONSUMPTION OF WOMEN EARNERS

Independent Variables	Dependent Variables: *Ln* of workplace-Residence Separation	*Ln* of House Size	*Ln* of House Amenities
Relative to married, no children			
Married with children	.073	.089[a]	−.056
Single with children	−.262[c]	.036	.052
Single, no children	−.343[a]	−.207[a]	.176[b]
Estimated wage	1.478[a]		
Residence			
within 5 miles of center	−.248[a]		
15–30 miles from center	.326[a]		
in city of over 500,000	−.161[b]		
Ln of job tenure	−.109[a]		
Ln of weekly hours	.086		
Black	.578[a]	−.026	−.221[a]
Moved in 1975	.140	−.051	.156[a]
Age	.019	.019[a]	.011
Age squared	−.0003	−.0001	−.0002[a]
Own home	.117	.172[a]	.057
Ln of household income		.074[a]	.286[a]
Estimated *ln* of worktrip		.072	.167[a]
Intercept	−.283	.205	5.213
MSE	.7215	.0773	.2555

[a]Prob > .99 that coefficient is nonzero.
[b]Prob > .95 that coefficient is nonzero.
[c]Prob > .90 that coefficient is nonzero.

these locational differences are significant. The household demographics are only significant when households trade off between size and other amenities in their housing choices. However, it is the presence of children rather than the presence of a second earner which effects housing choices. Households with children select larger homes with fewer other amenities, *ceteris paribus*.

Table 3 presents the estimated coefficients of the housing consumption model for women earners. Because there is no evidence that marriage is selective of women's earnings characteristics and because there are very few women earners in one-earner marriages, Table 3 includes both married and single women earners.

As in Table 2, the independent variables in Table 3 have the expected effects. However, in Table 3 the household demographic variables are more significant. Women earners in one-earner households (i.e., single with and without children) live significantly closer to their jobs than women earners in two-earner households. The tradeoff between size and other housing amenities is consistent with the result in Table 2. Larger households select larger houses with fewer other amenities; that is, the rank order of house-size effects is: married women with children; single women with children; married couples; single women.

The married men and women in this study are married to each other and, therefore, residing at the same locations. Tables 2 and 3 can be used to compare marrieds and singles as well as one-earner and two-earner married households. Table 2 indicates that there is no statistically significant behavioral difference in the household's choice of workplace-residence separation in one-earner and two-earner married households. Table 3 indicates that single women live significantly closer to their jobs than married women. Therefore it seems that the increasing labor force participation of married women should not change the urban analyst's standard expectation that

household size and income determine the proximity of the household to the urban center.

IV. Conclusions

There are four primary empirical findings in this study:

1. Two-earner, husband-wife households behave as one-earner, husband-wife households in their choice of housing location, house size, and house quality; that is, the observed differences are fully explained by fertility, income, and job characteristics.
2. Unmarried individuals live closer to their jobs, purchase smaller and higher quality housing units than married persons, *ceteris paribus*.
3. Households with children trade off house quality for larger housing units, *ceteris paribus*.
4. While married women earners with children reside in more suburbanized locations than other earners, their commuting behavior implies that their workplaces are substantially more suburbanized than that of other earners.

These results do not imply, however, that the growth in two-earner households will not significantly change urban form. Rather they imply that the residential land use effects of increased labor force participation of women will occur indirectly through the effects of that participation on household income and household size. If the increased labor force participation of women leads to higher age at first marriage, higher divorce and separation rates, and lower fertility, then it will decrease household size and increase demand for smaller, more centrally located housing. On the other hand, higher participation increases household money income, thus increasing demand for housing, and encouraging residential suburbanization. Increases in female labor force participation among married women with children also increase the incentive for employers to suburbanize. The net long-term effects of growth in the female labor force on urban land markets depends ultimately on the long-term demographic effects of female labor force participation.

REFERENCES

Edwin S. Mills, *Urban Economics*, Glenview 1972.

M. J. White, "A Model of Residential Location Choice and Commuting by Men and Women Workers," *J. Reg. Sci.*, No. 1, 1977, *17*, 41–52.

Consequences of the Rise of the Two-Earner Family: The Breakdown of the Sexual Division of Labor

By JULIE A. MATTHAEI*

The entrance of increasing numbers of married women into the labor force in the twentieth century, and the concommitant rise of the two-earner family, are well documented. This paper will analyze the meaning of these developments for American economic and social life by placing them within a conceptual and historical framework, arguing that they represent the breakdown of the sexual division of labor.

Neoclassical economic theory fails to provide a starting point for the analysis of the sexual division of labor and its transformation through history. The work experiences of the sexes are seen as sources of utility which do not otherwise affect the worker's self-constitution. The choices of the sexes within the household to specialize in different kinds of work, market or nonmarket, are "explained" as the result of the different preferences of the sexes. But since individual preferences are outside of the realm of economic theorizing, this difference in preferences can not itself be explained—neither can the possibly sexist preferences of employers or employees.

My analysis will begin with a different conceptual framework. Social activities, in particular "work," construct the individual worker's identity. The division of social activities between the sexes, or the "sexual division of labor," is the source of the masculine/feminine difference.[1]

Society constructs two different kinds of beings out of its young, men and women. A properly socialized female undertakes woman's work, shuns man's work, and consequently develops herself into a woman; a properly socialized male, likewise, becomes a man by undertaking man's work and avoiding woman's work. A biological female is not automatically a woman, nor is a male automatically a man. It is the difference of the social activities of the sexes which makes them into different social beings.

The sexual division of labor, while unconsciously produced by society, nevertheless has a "reason" behind it. It turns males and females into different and complementary beings who need each other to be whole, giving reason to their unity in marriage, and establishing reproduction as a social process (see Claude Levi-Strauss and David Levine).

It was in the nineteenth century, with urbanization and the development of factory wage labor, that the "traditional family" as we know it today became prevalent. Family life gained a distinct sphere of its own, the home, and the "cult of domesticity" recognized and elevated woman's previous duties of caring for the family. In particular, mothering was first acknowledged as a significant social activity. It became the center of woman's new homemaking vocation. Meanwhile more and more men gained property rights on the basis of which they competed for self-advancement in the labor force hierarchy as capitalists or wage workers.[2]

*Assistant professor of economics, Wellesley College.

[1]While this division of activities between the sexes has been termed by social scientists "the sexual division of labor," it is not properly speaking a division of labor, i.e., of wage labor, between males and females, but rather a division of social activities. Indeed, until recently, it assigned wage labor predominantly to men, giving to women responsibility for private work in the home.

[2]The limits of this article do not allow me to show how this "traditional family" emerged with the development of the factory system out of the colonial family economy. For a more detailed analysis of these historical transformations, as well as those in the twentieth century, see my forthcoming book.

In a marriage, man's work and woman's work were complementary. A homemaker's care of the family and maintenance of the home allowed her husband to focus all of his energies on his job. A husband's earnings in the labor force allowed his wife to dedicate herself to the private care of the home and family. Being the family provider was of central importance to manhood; indeed, for the majority of men, whose jobs offered little or no potential for self-advancement, manhood was centered in one's ability to support a homebound homemaker. Marriage was postponed and saved for until the husband's earnings could keep his wife at home. Children of both sexes were sent out after supplementary income, work was taken into the home, even charity sought, before the wife was sent into the labor force.

Hence the sexual division of labor was expressed in the economy in the different labor force participation patterns of the sexes. High labor force participation rates characterized adulthood for men, and adolescence or widowhood for women (see Table 1). The sexual division of labor was also at the root of the segregation of the sexes in the labor force into men's jobs and women's jobs.

Men and women, as constituted by the sexual division of labor, were different, and entered the labor force differently. A man sought enough wages to support a wife and family, and, if possible, to advance himself as an individual in the labor force hierarchy. He was willing to subordinate his life to the requirements of his job in order to gain advancement; his marriage to a homemaker both required and enabled him to do so.

A woman, as a present or future homemaker, sought to satisfy the particular and changing needs of her family, as a daughter, wife, and/or mother. If she entered into the labor force it was not to seek self-constitution or a life time vocation but simply money to help fill commodity needs of her family which, for some reason, were unfilled. Success for the wage-earning woman was not advancing herself in the labor force, but being able to leave the labor force to dedicate herself full time to homemaking.

TABLE 1—LABOR FORCE PARTICIPATION RATES BY SEX AND MARITAL STATUS, 1890
(Population 10 Years of Age and Over)

	Males	Females
Total	77.29	16.97
Single and Unknown	60.08	28.16
Married	95.93	4.63
Widowed	79.52	29.25
Divorced	90.80	48.99

Source: U.S. Bureau of the Census, 1890.

Faced with these two different kinds of labor, the developing system of capitalist firms established sex-typed jobs; each particular job was constituted, from its inception, either as a man's job or as a woman's job. Jobs requiring the full-time, full-life commitment of an aggressive and self-advancing individual were established by firms to tap the energies of men. Women were neither interested in nor suited for such jobs. On the other hand, because they merely sought supplementary wages to aid their families, women were willing to accept part-time, seasonal, and servile jobs. These became women's jobs: all but the most oppressed men refused such jobs, for they failed to offer the financial rewards or social recognition required for manhood.

Interestingly, even jobs which did not exclude either women or men, such as unskilled factory work, were sex-typed at least by establishment. If low-wage women were available and employed, this low wage, combined with the stigma of doing woman's work, would exclude men. If men were employed, they demanded a family wage and fought to exclude woman's low-wage competition (see Heidi Hartmann). And the stigma of doing man's work kept most women from trying.

Hence the sexual division of labor in marriage was reflected into the economy first in sex differences in labor force participation patterns, and secondly in the existence of man's jobs and woman's jobs. The presence of a minority of women in the labor force reinforced rather than undermined the ideal of domestic womanhood. For a woman, labor force participation

clearly represented failure—either her own failure as a woman to find and keep a manly husband, or her husband's or father's failure as a man to support her financially. The segregation of women into an inferior set of women's jobs further "proved" woman's unfitness for the masculine economic world.

Twentieth-century developments have drawn married women into man's sphere of the labor force, and in some cases, into men's jobs. The maturation of capitalism in the late nineteenth and early twentieth centuries brought a tremendous expansion in the production of consumer goods. The development of a market for these involved the establishment of an expanding set of needs which could only be filled by these commodities. As this system of needs was developed and universalized, a common idea of a minimum "standard of living" emerged, delineating the commodity requirements of individual and family life. Homemaking, the filling of the particular needs of the family, now necessitated providing the family with these needed commodities, commodities for which homemade efforts could provide no real substitute. These commodities freed the homemaker from much of the drudgery involved in housework; now she filled many of the family's needs by participating in the economy as a consumer of commodities.

In the middle-class family, the homemaker filled the family's commodity needs through consumption alone, spending the income of her husband. However, if a homemaker's husband's income could not provide her family with the commodities she deemed necessary, the requirements of homemaking became unclear and contradictory. The family's needs for extra income, filled by the homemaker through her own labor force participation, conflicted with her husband's and family's needs for her presence in the home. With the advance of the twentieth century, the expansion of commodity production and the growing belief that all families deserve the standard of living have eroded the domestic ideal in many working class families, sending increasing numbers of homemakers into the labor force to fill expanding family needs (see Table 2).

At the same time, the very women who have been allowed by their husbands' incomes to realize the domestic ideal of homemaking have become increasingly dissatisfied with it. The development of homemaking into a vocation in the nineteenth century brought recognition of women as responsible and important social beings. Woman's higher education became desired training for the homemaking vocation, and, in particular, for effective mothering. But education qualified woman not only for homemaking, but also for careers outside of the home, in man's sphere. Furthermore, while effective mothering has increasingly required the education and individuality of the homemaker, its practice cannot provide her with either the public and financial recognition or the opportunity for self-development necessary to sustain that individuality.

The contradictory demands of the homemaking vocation have propelled increasing numbers of women out of the home. Many of these women have sought self-development and recognition in careers in the labor force. Yet they have discovered that most careers are monopolized by men, and that employers often refuse to consider females for men's jobs, regardless of their qualifications. The attempts of these women to break into men's jobs, and the difficulties they have encountered in doing so, have brought the sex-typing of jobs into the spotlight. The historical exclusion of women from highly paid, skilled, professional, and management jobs, and their concentration in low-paid low-status ones have been discovered and decried as contrary to the American principle of equal opportunity.

With the progress of the twentieth century, the two-earner marriage has become more prevalent than the husband wage earner/wife homemaker marriage. An active feminist movement has attacked sex discrimination and inequality between the sexes; the government has responded with nondiscrimination policies, including the de-sexing of job titles. The stigma against females doing man's jobs is breaking down,

TABLE 2—MARITAL STATUS OF WOMEN IN URBAN LABOR FORCE, 1890–1978
(Persons 15 Years Old and Over, 1890–1930; 14 Years Old and Over, 1940–66; 16 Years Old and Over, Thereafter)

	Percent Distribution of Female Labor Force				Labor Force Participation Rates				
	Single	Married Total	Married Husband Present	Widowed or Divorced	Total	Single	Married Total	Married Husband Present	Widowed or Divorced
1977	24.1	61.3	56.8	14.8	48.0	58.9	47.1	46.6	39.0
1970	22.5	62.3	57.1	15.0	41.6	50.9	40.2	39.6	36.8
1960	23.6	60.7	55.2	15.7	34.5	42.9	31.7	30.6	36.1
1950	31.9	52.2	46.5	16.0	29.0	46.3	23.0	21.6	32.7
1940	49.0	35.9	30.1	15.0	25.8	45.5	15.6	13.8	30.2
1930	53.9	28.9		17.2	24.8	50.3	11.7		34.4
1920	77.0	23.0		[a]	23.7	46.4	9.0		[a]
1910[b]	60.2	24.7		15.0	25.4	51.1	10.7		34.1
1900	66.2	15.4		18.4	20.6	43.5	5.6		32.5
1890	68.2	13.9		17.9	18.9	40.5	4.6		29.9

Sources: U.S. Bureau of the Census: for 1890–1970 (see 1971, p.133); for 1978 (see 1978, p.404).
[a]Single includes widowed or divorced
[b]Data not comparable: census enumerators encouraged to record "the occupation, if any, followed by a child of any age or a woman" included more women in labor force.

and women on all levels are beginning to seek entrance into the more lucrative and more respected man's jobs.

These developments represent major transformations in the sexual division of labor, including direct attacks on its expression in the labor force, the sex-typing of jobs. What now proves the most surprising is the tenacity of the sex-typing of jobs in spite of these attacks. Studies show that the level of occupational segregation of the sexes has not changed substantially since 1900! (See Edward Gross; Francine Blau and Wallace Hendricks.)

The rise of the two-earner family does not in itself represent the disintegration of the sexual division of labor. Working-class homemakers are entering the labor force as an extension of their homemaking work; hence they enter woman's jobs, jobs compatible with their homemaking commitments. Married women seeking "second careers" in the labor force attempt to combine their homemaking vocation with a man's career. These women are not rejecting homemaking and traditional womanhood, but simply trying to extend or supplement it. Little wonder that most women face failure in their attempts to succeed in men's careers (see M. Henning and A. Jardim; C. F. Epstein).

Nevertheless these developments are slowly contributing to the breakdown of the sexual division in marriage and in the labor force. First, the two-earner family is less stable than the husband-provider/wife-homemaker family. The wife's labor force earnings make her less financially dependent upon her husband, and make him less responsible for her well-being. Meanwhile, since the wage-earning wife shares with her husband the "job" of supporting the family financially, there is pressure on him to share her homemaking responsibilities, pressures which are intensified if she is working in a man's job. Conflicts over respective roles develop; divorce and remarriage are possible "solutions."

As the basis of the traditional marriage is eroded, a new conception of marriage is emerging, the "companionate marriage."[3] As females take on masculine as well as feminine activities, becoming more whole human beings, they urge males to experi-

[3]The emergence of a new, companionate marriage has been noted by sociologists and psychologists, although they have difficulty realizing that this change means the end of masculinity and femininity.

ence their feminine sides. Males are drawn into active participation in family life, including child rearing, while at the same time allowed to give up their total responsibility for heading and financing the family. The new marriage is a union, not of man and woman, but of full persons, each of whom participates actively in economic and family life.

However the emergence of the new companionate marriage to replace the disintegrating traditional marriage is blocked by the latter's expression in the labor force—the sex-typing of jobs. As we have seen, sex-typed jobs have been constructed to fit man and woman, as different and complementary social beings constructed by and sustained in the traditional family. The full individual emerging today within the companionate marriage cannot find fulfillment in such a sex-typed job structure. Men's jobs, which allow the development of individuality, preclude an active family life; women's jobs, while they are compatible with an active family life, do not allow the development of individuality. The breakdown of the sexual division of labor in marriage and its reconstruction as a marriage of equal and similar persons, then, are both necessary for and contingent upon the destruction of the sex-typing of jobs and the emergence of a new job structure.

The rise of the two-earner family represents an early stage of the breakdown of the sexual division of labor, part of a process from which there is no turning back. The particular manner in which these transformations will take place is not yet clear; indeed the changes in social life involved are so fundamental as to still be virtually inconceivable. I question whether capitalism can survive without "men" willing to dedicate themselves to its service in exchange for self-advancement, at the price of all else; it seems certain that if society is to achieve the construction of the full person and the new family it will have to self-consciously reconstruct economic life.

REFERENCES

F. Blau and W. Hendricks, "Occupational Segregation by Sex: Trends and Prospects," *J. Human Res.*, Spring, 1979, *14*, 197–210.

C. F. Epstein, *Woman's Place: Options and Limits in Professional Careers*, Berkeley 1970.

E. Gross, "Plus Ca Change...the Sexual Structure of Occupations Over Time," *Soc. Problems*, Fall 1968, *16*, 198–208.

H. Hartmann, "Capitalism, Patriarchy, and Job Segregation by Sex," *Signs*, Spring 1976.

Margaret Hennig and Ann Jardim, *The Managerial Woman*, Garden City 1977.

D. P. Levine, "*Groundwork for a Theoretical Treatment of the Family*," New Haven 1975.

C. Levi-Strauss, "The Family," in Arlene Skolnick and Jerome H. Skolnick, eds., *The Family in Transition*, Boston, 1971.

Julie Matthaei, *Womanhood and Economic Life: A Study in American Economic History*, New York forthcoming.

Mary P. Ryan, *Womanhood in America, from Colonial Times to the Present*, New York 1975.

U.S. Bureau of the Census, *Eleventh Census of the United States*, Vol. 9, Part 2, "Population," Washington 1890.

______, *Historical Statistics of the United States; Colonial Times to 1970*, Washington 1971.

______, *Statistical Abstract of the United States 1978*, Washington 1978.

Real Income Equivalence Among One-Earner and Two-Earner Families

By EDWARD P. LAZEAR AND ROBERT T. MICHAEL*

As the labor force participation rate of married women climbs above the 50 percent mark, there is growing interest in the impact of the two-earner family on the level and distribution of family income. The popular press details the affluence of "America's New Elite" and researchers investigate the distribution of family income (see Richard Layard and A. Zabalza for the United Kingdom and James Smith for the United States). Recent *Current Population Survey* (*CPS*) reports indicate that wives now contribute about 18 percent of the family's before-tax income.

In order to make comparisons across households, however, it is important to understand the mapping from dollar income into some notion of "real income" and how that mapping is affected when one moves from one-earner to two-earner families. In an earlier paper, we proposed a technique for making income comparisons across various household sizes. In this paper, we employ a modified version of that method to convert nominal income to some real equivalent comparable across families of a given size, but with one or two family earners. We begin with a discussion of some of the conceptual difficulties involved in making cross-family size or structure comparisons in income. We then describe spending pattern differences between one- and two-earner families. The final section discusses our effort to convert nominal income into comparable real income units. We conclude that nominal income differences between two- and one-earner families far overstate differences in standards of living.

*University of Chicago and National Bureau of Economic Research; Stanford University and National Bureau of Economic Research, respectively. This research has been supported by grants from the Alfred P. Sloan Foundation and the National Science Foundation. We acknowledge the research assistance of Wendy Max and helpful comments from Tom MaCurdy and Victor Fuchs.

I. Some Theoretical Issues

We define the family's standard of living to be the market value of the bundle of service flows consumed by the family. Conceptually the bundle is an objectively measurable entity which, when priced at some observable cost per item, yields a value appropriate for comparison across families. It is analogous to the money cost of a given bundle of market goods and services, but in principle captures differences in family technology in transforming market goods into service flows.

The transformation of money income into service flows probably differs between one-earner and two-earner families for several reasons, including 1) income taxes, 2) costs of employment, and 3) differences in household techniques. Unless one thinks the service flows from government services are proportional to taxes paid, differences in after-tax income rather than before-tax income more adequately reflect differences in standards of living. Also, there are certain costs of employment, such as costs of transportation to work, requisite clothing expenditures for employed persons, union or professional fees, etc. These taxes and work-related costs should be netted out of income in obtaining an indicator of standard of living. More elusive but potentially large differences in standards of living may result from differences in production technologies used by different family structures to achieve a given service flow. Two-earner families presumably use some market services to perform tasks typically performed by the not-employed spouse in the one-earner family. This choice of technology—say for producing a year's supply of

clean, ironed shirts—affects whether a substantial dollar expenditure or a trivial dollar expenditure is required to achieve the same real service flow. The differences in money income and expenditure that reflect different technologies of production but not different levels of service flow do not reflect differences in standards of living.

We argue, therefore, that it is desirable to convert money income into a standard of living measure by adjusting for differences in taxes, work-related costs, and household technologies. It has not yet proven feasible to estimate the household technologies directly because the output from the production processes are generally not observed. In fact one usually has measures of only one set of productive inputs, the purchased market goods and services. Our technique, outlined below, allows us to begin to take account of differences in household technologies and to convert nominal income into more comparable equivalent income figures.

II. Differences in Measured Income and Spending Patterns

From the 1972–73 U.S. Bureau of Labor Statistics (*BLS*) *Consumer Expenditure Survey* we selected two groups of husband-wife families *with no children* with male under age 65: 1) "Two-earner families" included the 869 families in which both the husband and wife were employed full time (i.e., usually worked 35 hours or more per week and worked more than 26 weeks during the survey year); 2) "One-earner families" included the 717 families in which the husband worked full time and the wife was not employed in the survey year.

If one looked at the average before-tax income of wives in two-earner families ($6,668) and compared that to the average before-tax income of their husbands ($10,581), one might conclude that these women raise their family's income by over 60 percent. More reasonable, as Table 1 shows, comparing the two-earner families' average before-tax family income ($17,249) to that of the one-earner family ($12,803), we find a 35 percent income differential. In after-tax family income, however, the differential is only 25 percent. Adjusting by regression for the differences in characteristics in these two groups (i.e., adjusting for the higher schooling levels and younger ages in two-earner families, as well as for trivial differences in region, city size, rurality, color, and survey year) the after-tax family income of the two-earner family is only about 20 percent higher.

The *BLS* data are useful in describing why the two-earner families had only $2,758 more after-tax income (unadjusted): the wives' average before-tax income was $6,653, but the increment in family income tax was $1,683 and the decrement in reported before-tax income of husbands was $2,212. This 20 percent income differential may seem small, but is roughly comparable to the differentials reported from the 1978 *CPS*.[1] So despite the impression in the popular press that two-earner families have very high incomes, evidence suggests that unadjusted or adjusted for several family characteristics, the differential in family income is only about 20 or 25 percent.

Turning to differences in spending patterns, we find the differential in total current consumption to be about 17 percent. For the seven components that exhaust total consumption, we find relatively big increases by two-earner families on expenditures on clothing (54 percent), durables (45 percent), and transportation (32 percent) and relatively little increases in the spending on food (2 percent) and services (4 percent). The big increase on clothing presumably reflects the wife's clothing expenditures for employment-related clothing needs (see the detailed items in Table 1), and the big increase in durables and transportation expenses are probably related to employment-related automobile expenditures as well (note also the sizable increase in gas

[1]Before-tax family income differentials for families grouped by generally analogous definitions of worker status (full time for husband, full time or not in the labor force for wife, but for families of all sizes) show differentials of 37 percent for young family heads (18–24) declining to differentials of only about 13 percent for groups defined by family head's age 45–54.

TABLE 1—INCOME AND EXPENDITURES BY ONE-EARNER AND TWO-EARNER FAMILIES OF SIZE TWO

	Two-Earner Families	One-Earner Families	Percentage Difference Adjustment[a]		
			Actual	1	2
Income Before Taxes	$17,249	$12,803	35		
Income After Taxes	13,884	11,126	25	22	20
Expenditure Items and Selected Subitems					
Total Current Consumption	$10,830	$9,236	17	10	7
Food	1,854	1,815	2	8	8
At home	1,088	1,285	−15		
At restaurants	529	341	55		
Clothing	678	439	54	30	24
Women's clothing	408	255	60		
Housing (renters only)	2,533	2,216	14	2	2
Rent	1,951	1,743	12		
Fuel	253	208	22		
Domestic service	57	41	39		
Transportation	662	503	32	16	13
Durables	2,239	1,545	45	17	15
Vehicle	1,166	772	51		
House furnishings and equipment	566	434	30		
Nondurable Goods	701	635	10	9	3
Gasoline	437	371	18		
Services	2,163	2,080	4	9	−1
Health insurance	264	263	0		
Life insurance	272	272	0		
Health care	234	330	−29		
Personal care	143	130	10		
Dry cleaning	109	77	42		
Gifts	585	612	−4		

[a]Adjustment 1 uses coefficients estimated on the two-earner families and the characteristics of the one-earner families to predict the income or expenditures the one-earner families would have made if they had had two earners. Adjustment 2 uses coefficients estimated on the one-earner families with the characteristics of the two-earner families to predict the income or expenditures the two-earner families would have made if they had had one earner.

expenditure). The small change in food expenditure masks the shift from grocery to restaurant expenditure seen in the detailed data. Likewise, the small change for services reflects the offsetting reduction in direct health care expenditures and the substantial increase in other service costs such as dry cleaning. Notice, too, the large increase in domestic services. Adjusting these comparisons for differences between the two subsamples in age, schooling level, region, city size, color, rurality, and year of survey, the clothing, transport, and durables expenditure differences remain relatively high while services, housing (rental units only), and nondurables remain relatively low. The average itself is a quite modest 8 percent increase in total current consumption.

This descriptive analysis of income and spending differences indicates that adjusted for socioeconomic characteristics, two-earner families had about 20 percent higher income after taxes and spent about 8 percent more in total expenditure.[2] That added expenditure went predominantly for items

[2]Implicitly, the two-earner families have substantially higher savings rates. Perhaps this reflects a short-term labor force attachment by the second employed spouse. If their employment were intermittent, we might expect the two-earner families to save during periods with two earners in order to smooth their consumption flows over time.

related to the additional employment and for items that essentially substitute market-produced services for traditionally home-produced services. How much of the added income or expenditure in two-earner families constitutes an increase in the family's standard of living? In the following section we illustrate an approach that we believe useful in converting dollars of expenditures on goods into their corresponding service flows.

III. Estimation of Income Equivalence

Comparing income in "real" terms across families with employed and nonemployed wives has been done by I. Serigeldin and Reuben Gronau, among others. Their strategies directly evaluate the cost of home-produced commodities, and specifically take account of the use of the wife's time in the nonmarket sector. Our strategy is to infer from market expenditures the information necessary to express money income in comparable units across families of different types. This revealed preference approach has been used in other studies as well to infer equivalence in income across different family sizes. (See John Muellbauer, A. P. Barten, and A. M. Henderson for examples.) Our procedure is derived from our earlier paper. The approach assumes that one- and two-earner households have the same underlying demand system for services. Differences in expenditures occur because the transformation of goods into services varies with the way in which one's time is spent.[3]

Our method converts income to comparable service flows across households of different size or structure. It involves knowing the price elasticities of demand and differences in expenditure patterns by family type. From this, we can infer the relationship between services per dollar in one family structure compared to another. These rates of transformation of dollars into service flows are used to calculate the equivalent levels of income that yield the same potential flow of services across family types.

More specifically, if the flow of services of type i, S_i is defined to be equal to the amount of market good X_i consumed in a two-earner family, then the price of a unit of S_i is $P_{si}=P_{xi}$ and the demand for S_i presumably is responsive to changes in P_{si}. Now let the technology which transforms goods into services in the one-earner family differ (potentially) from that of the two-earner family. Parameterize it so that in a one-earner family, $S_i=X_i(1+J_i)$ as opposed to the $S_i=X_i$ applicable in the two-earner family. If we characterize demand by an ordinary linear expenditure system derived from a Stone-Geary utility function for the two-person family,[4] for the two-earner family, the derived demand for X can be written

$$(1)\qquad P_{x_i}X_i=\gamma_i P_{x_i}+\beta_i\left(Y-\sum_{j=1}^{n}\gamma_j P_{x_j}\right)\qquad i=1,\dots,n$$

while for the one-earner family, assuming no differences in the demand structure:

$$(2)\qquad (P_{x_i}X_i)'=\gamma_i\frac{P_{x_i}}{(1+J_i)}+\beta_i\left(Y'-\sum_j \gamma_j\frac{P_{x_j}}{1+J_j}\right)$$

Differencing equations (1) and (2) gives

$$(3)\qquad (P_{x_i}X_i)'-(P_{x_i}X_i)=\gamma_i P_{x_i}\left(\frac{1}{1+J_i}-1\right)+\beta_i(Y'-Y)-\beta_i\sum_j\gamma_j P_{x_j}\left(\frac{1}{1+J_j}-1\right)$$

The research strategy involves estimating the J_i for each consumption item by using known (prior estimated) values for the

[3]Other things constant, then, differences in labor force participation across households arise from different random draws from the wage offer distribution or from the reservation wage distribution.

[4]The Stone-Geary is employed merely for computational convenience.

parameters of the demand system γ and β (from Michael Abbott and Orley Ashenfelter), observed prices P_x assumed constant for all households at a point in time, observed values for $(P_{xi}X_i)'$ and Y', and values for $(P_{xi}X_i)$ and Y predicted from estimates calculated for the two-earner families and applied to the demographic characteristics of the one-earner families. The predicted $(P_{xi}X_i)$ and Y indicate what each one-earner family would have spent on X_i (i.e., food or clothing) and what its after-tax family income would have been if it had been a two-earner family. With this, we have observed values for everything in equation (3) except $J_1,\ldots,J_i,\ldots,J_n$. But we have a separate analogous equation for all n items. Hence we have a system of n equations in n unknowns.

The observations are a subset of 180 of the 717 one-earner families whose expenditures were described in the previous section. The 180 families are renters with head $\leqslant 35$-years of age. Families with older heads would include some which previously had children and whose expenditure patterns may reflect that fact. Table 2 shows their average actual expenditures and the average predicted hypothetical expenditures if these couples had been two-earner families. From these, the J_i are solved for, using the mean values of expenditures and income. Savings is neglected in this framework because we have been unable to deal with this category appropriately either theoretically or empirically. Also, durables are excluded from the analysis for two reasons. First, they are much like savings and our failure for savings extends as well to durables. Second, durable expenditures are lumpy which creates problems of aggregation and also affects the precision of expenditure system estimates of the durable good parameters. At present, we believe cleaner estimates are obtained by deleting the category rather than forcing our method to deal with durables. Of course, this means our estimates of equivalence apply, at best, to only a subset (about 80 percent of total current consumption) of the service flows that constitute the standard of living.

TABLE 2—ACTUAL AND HYPOTHETICAL EXPENDITURES BY ONE-EARNER FAMILIES
(Renters Only, Head Aged $\leqslant$ 35 Years: 180 Observations)

	Actual	Hypothetical	J_i
Food	\$1032	\$1325	.48
Clothing	445	626	1.21
Housing	2326	2334	−.03
Transportation	542	620	.86
Nondurables	567	663	.24
Services	1040	1166	.19
Total Expenditures	\$5952	\$6735	

The weighted average of these J_i is 0.30 suggesting that the average price of a unit of services in the one-earner family, relative to its price in the two-earner family, is $1/1+\bar{J} = 1/1.3 = 0.77$. Service flows are only 77 percent as expensive in one-earner families probably because more time is combined with goods to produce services there. The differences are most pronounced for clothing, transportation, and food.

Said differently, a two-earner family spending \$1,000 on market goods achieves the same level of service flow as a one-earner family spending $\$1000/(1+\bar{J}) = \$770$. So in service flow units, the \$770 expenditure in the one-earner family is equivalent to the \$1,000 expenditure in the two-earner family. Or, the actual expenditure of \$5,952 on these items by these one-earner families is equivalent to an expenditure of \$7,730 in a comparable two-earner family.

IV. Evaluation and Summary

We find that two-earner families have about 20 percent more money income after tax than other comparable one-earner families, but our rough estimate of the differences in technologies in the home suggests that the average two-earner family requires about 30 percent more money income to achieve the same level of standard of living as a one-earner family, for the subset of income spent on nondurable goods and services. So from our evaluation the two-earner family does not appear to have a higher standard of living.

Three caveats are in order:

1) We have looked only at one particular type of family in constructing an equivalence measure, namely renters, age 35 and under, in families with no children, and we report results which pertain to a subset of expenditures which excludes durables and ignores savings.

2) Findings in Table 1 reveal that the composition of the composite commodities changes as we look across family types; future work should take that into account. The advantage of the usual definitions of composite commodities is that they allow us to use estimates of utility parameters from the literature. However, failure to take account of changes in the composition of the composite goods as family type changes means that our estimates of *J*s are biased. In subsequent work one might want to restructure the composite commodities in a manner more consistent with the definition. That is, weights of items within the composite good should be approximately consistent across family types.

3) We do not include any valuation of the psychological or sociological advantages of one family structure over the other. Our procedure for holding constant other arguments in the utility function is admittedly deficient.

We do not encourage the reader to think our estimates more than illustrative. But the issue addressed here is an important one for measurement of economic performance, and for issues of cross-family equity in social policymaking. We believe that some progress has been made on this very difficult subject.

REFERENCES

M. Abbott and O. Ashenfelter, "Labour Supply, Commodity Demand and the Allocation of Time," *Rev. Econ. Stud.,* No. 3, 1976, *43*, 389–411.

A. P. Barten, "Family Composition, Prices and Expenditure Patterns," in P. E. Hart, et al., eds., *Econometric Analysis for National Economic Planning*, London 1964.

R. Gronau, "The Intra-Family Allocation of Time: The Value of Housewives' Time," *Amer. Econ. Rev.*, Sept. 1973, *63*, 634–51.

A. M. Henderson, "The Cost of Children, Part I," *Pop. Stud.*, Sept. 1949, *3*, 130–50; "Parts II, III," Dec. 1950, *4*, 267–98.

R. Layard and A. Zabalza, "Family Income Distribution: Explanation and Policy Evaluation," *J. Polit. Econ.*, Oct. 1979, *87*, Part II, S133–62.

E. P. Lazear and R. T. Michael, "Family Size and the Distribution of Real Per Capita Income," *Amer. Econ. Rev.*, Mar. 1980, *70*, 91–107.

J. Muellbauer, "Testing the Barten Model of Household Composition Effects and the Cost of Children," *Econ. J.*, Sept. 1977, *87*, 460–87.

I. Serigeldin, *Non-Market Components of National Income*, Ann Arbor 1969.

J. P. Smith, "The Distribution of Family Earnings," *J. Polit. Econ.*, Oct. 1979, *87*, Part II, S163–92.

U.S. Bureau of Labor Statistics, *Consumer Expenditure Survey*, unpublished data (on tape).

Labor Force Participation Rates of Women and the Rise of the Two-Earner Family

By DANIEL C. QUINLAN AND JEAN A. SHACKELFORD*

By the end of the first quarter of 1978, the typical husband-wife family in the United States consisted of two wage earners. While only 30 percent of female workers in 1930 were married, today 55.6 percent of all female workers are married with husband present. Moreover, this growth in the labor force participation of women has occurred despite the persistence of wage inequalities and occupational stereotyping in the workplace. Women continue to be segregated in clerical, sales, and service occupations, and to receive substantially lower salaries than their male counterparts. The immediate problem then is to explain why married women, commonly designated as the least-expected subgroup of women to enter the labor force, are exhibiting the greatest increase in participation rates when the labor market is still primarily a male domain.

In this report, following Valerie Oppenheimer, we break away from traditional models of female labor force participation that normally emphasize supply factors and instead propose a model that stresses demand factors in explaining the tremendous growth in gainfully employed married women. The inadequacies of supply-oriented models have been admirably noted elsewhere: consequently, we move directly into an elaboration of the demand model.

A number of interrelated factors have contributed to the creation of increased demand for female labor. First, since 1900, the American economy has shifted away from primary industries toward secondary and tertiary industries in terms of the proportion of people employed. For women, however, there has been only a slight distributional shift toward the tertiary sector. Women, therefore, have increased their labor force participation in precisely those industries where they have historically been located, that is, nonprimary industries. There appears to be then no direct link between interindustry changes and the rise in female labor force participation. The sectoral shifts for men, however, are quite different. Men have shifted away from the primary sector and have moved into the secondary and tertiary sectors. Generally, the secondary and the tertiary sectors tend to employ a higher proportion of managerial and professional personnel. Furthermore, across practically all industries, there has been a pronounced increase in the proportions of these two types of occupations. Between 1950 and 1970, the percentage of men in managerial and professional jobs more than doubled. The connection between changes in the male occupational structure, sectoral shifts in the economy, and the increased labor force participation of women, now becomes evident. It is the movement of the *male* occupational structure, generated in part by sectoral changes and the rise of corporate and governmental influence on the economy, toward an increasing proportion of managerial and professional workers that has to a large extent stimulated the rise in female employment. Since these occupations normally require adjunct clerical personnel, which is a sex-stereotyped aggregate of occupations, the demand for this specific type of female labor has greatly expanded. It is no wonder therefore that the type of occupations among women which has demonstrated the largest proportionate increase since 1950 is the clerical occupation.

Yet, a change in the occupational composition of male workers cannot wholly account for increased female labor force participation since naturally not all women are secretaries. The rapid growth of the service sector has also imparted further demands for female workers, though there has been a movement away from domestic service to health and professional services.

*Bucknell University.

TABLE 1—ESTIMATES OF *LFP* FOR SELECTIVE MARITAL AND AGE GROUPS OF WOMEN

	$PCLER_{56}$	LFP_{56}	$PMAN_{56}$	$PSALES_{56}$	$PSERV_{56}$	R^2
Single Women						
16–19	−.091*	.825	.237	.204	−.132	.55
	.935**	101.927	6.856	1.856	.827	
20–24	.133	.239	−.174	−.175	.176	.04
	1.089	4.490	1.869	1.869	.7187	
25–29	.004	.247	.268	.141	−.228	.12
	.102	6.047	5.109	.478	1.282	
30–34	−1.310	8.370	.1603	−.007	.155	.01
	.011	.656	1.603	.108	.534	
35–44	−1.185	.005	.310	.168	−.257	.06
	.009	.285	6.389	.643	1.578	
45–65	.463	.131	−.004	−.200	.008	.18
	14.237	1.700	.108	1.035	.164	
Married Women						
16–19	.329	.441	−.365	−.225	.302	.37
	9.761	24.897	13.128	1.714	3.197	
20–24	.288	.607	−.480	−.408	.546	.62
	10.615	61.010	38.159	9.168	15.858	
25–29	.008	.823	−.201	−.298	.323	.79
	1.736	252.659	11.317	9.078	10.967	
30–34	.004	.795	−5.308	−.166	.165	.67
	.011	.659	1.602	.108	.534	
35–44	.143	.007	−.211	.190	7.719	.03
	1.329	.484	2.865	.792	.138	
45–65	−.006	−.001	−.003	−.002	−.002	−.04
	.230	.002	.052	.005	.008	
Other Women						
16–19	.224	.208	−354	.004	−.003	.04
	3.226	3.906	7.341	.034	.026	
20–24	.006	.150	.004	−.192	.162	.00
	.232	1.770	.069	.776	.588	
25–29	−.139	.222	.252	−.442	.405	.10
	1.319	4.968	4.168	4.600	4.088	
30–34	−.237	.154	.464	−.387	.422	.19
	4.332	2.508	15.037	3.953	4.967	
35–44	−.253	.008	.593	−.431	.406	.25
	5.058	.598	24.834	5.309	4.868	
45–65	−.005	−.003	.168	−.114	4.991	−.03
	.131	.071	1.647	.271	.054	

*unstandardized coefficient **F-statistic

In addition, growth in the tertiary sector has engendered increases in the only professional occupations historically female—teaching, nursing and health related personnel, and social workers. Clerical, teaching, medical, and service jobs accounted for 70 percent of all female jobs in 1974.

Thus, it is argued that occupational and sectoral changes in the American economy have generated a much greater demand for female labor and that this demand far exceeds the supply coming from the traditional sources of female workers—single women, women with grown children or no children, and women whose husbands are either absent, deceased, or divorced. Over time then, this increased demand has incrementally exhausted the more desirable pools of female labor to the point where women the least expected to enter the labor force (married women, husband present, with children under age 6) are now showing the greatest jump in labor force participation. Hence, women have responded favorably to this increased demand even though they still face considerable wage disparities and occupational segregation.

To test the power of this argument, we utilized the breakdowns by state of the U.S. Bureau of the Census reports, *Characteris-*

tics of the Population, which provides enumerations for occupations within general types of industries for each sex and labor force enumerations by marital status, age, and race. Attention is confined to the censuses of 1950, 1960, and 1970, and to the white population only. The attraction of this data is that it permits a disaggregation of *U.S.* trends in female labor force participation at least to the state level (firms would be preferred), and additionally the opportunity to observe temporal trends in the male occupational structure by industry and in the labor force status of specific pools of women labor. Thus, this data set incorporates both cross-sectional variation and variation over a brief period of time, thereby yielding the classic panel approach to social dynamics.

The design of the analysis follows the pooled cross-section method wherein 1950 and 1960 data are used to predict the state of female labor force participation in 1960 and 1970. The variables chosen to predict the labor force participation of selective age and marital status groups of women include the proportion of managers and professionals to all men 16 years and older gainfully employed in the state, the proportion of female clerical workers to all employed women 16 years and older, the proportion of female service workers, the proportion of female sales workers, and a lagged version of specific female labor force participation rates. The following equation summarizes both the design of the analysis and the models to be estimated:

$$\begin{aligned} LFP_{67} = a + b_1 LFP_{56} + b_2 PMAN_{56} \\ + b_3 PCLER_{56} + b_4 PSALES_{56} \\ + b_5 PSERV_{56} + u \end{aligned}$$

where LFP_{67} represents the labor force participation rates for six age categories of women (16–19, 20–24, 25–29, 30–34, 35–44, 45+) and three marital statuses (single, married-husband present, and other) in 1960 and 1970, where LFP_{56} is the same construct as LFP_{67} but for the years 1950 and 1960, where $PMAN_{56}$ denotes the proportion of managers in 1950 and 1960, where $PCLER_{56}$ is the proportion of female clerical personnel, where $PSALES_{56}$ designates the proportion of female sales personnel, and where $PSERV_{56}$ represents the proportion of service workers. Ordinary least squares is used to estimate this model although on the surface it may seem that a generalized least squares approach would yield better results. However, since the weight of the cross-sectional cases to the cross-panel cases is 49:2, it is extremely unlikely that the estimates will be unduly biased.

The results of performing these eighteen separate regressions are reported in Table 1. Clearly, they lend little support to the demand model; as a matter of fact, they generally contradict it. Looking at the six age groups of the marital category Other Women, it is apparent that none of the independent variables, including the lagged participation rates, exhibit any substantial and consistent power to predict the labor force participation of other women in 1960 and 1970. Much of the same can be said for the age groupings under the category Single Women. Only for single women, aged 16–19, does the lagged participation rate show any capacity to predict later participation. None of the remaining independent variables ($PMAN_{56}$, $PCLER_{56}$, $PSERV_{56}$, and $PSALES_{56}$) seem to exert any considerable impact. Furthermore, the signs of the coefficients are quite frequently in the wrong direction, that of predicting reduced participation rates. While it was expected that these remaining independent measures would possess little ability to forecast participation for these two categories of women given the relative temporal stability of their participation, it was not expected that the lagged measures would also be only weakly associated with later participation.

The results for the Married Women group are even more disappointing. Though for the first four age categories the adjusted R^2s are fairly large, the coefficients for $PMAN_{56}$ and $PSALES_{56}$ are negative and thus in direct conflict with the model's expectations. Part of the failure of $PMAN_{56}$ to show any sizeable effects could be due to the inclusion

of $PCLER_{56}$ in the equations with which $PMAN_{56}$ is highly correlated (approximately 0.74). Yet, when $PCLER_{56}$ is omitted, the coefficients for $PMAN_{56}$ are not significantly altered. The only notable results from this analysis are the first and second coefficients for $PCLER_{56}$ which are fairly substantial and in the correct direction. Married women between the ages of 16 and 24 do constitute a decent share of the married women labor pool. Unfortunately, the remaining coefficients are trivial and it is these latter groups of married women who are exhibiting the greatest increase in labor force participation.

By and large, these panel analyses offer little if any confirmatory evidence for the demand model. However, a few points should be made concerning these results. First, a number of extraneous influences have not been controlled in this analysis. It could be the case that female labor force participation is markedly affected by the age structure of states. Unusually elderly populations could cause a severe demand for the labor of married women since the supply of single women would likely be exhausted. The racial and ethnic composition of a state could also distort the labor force participation of white women. If female members of ethnic and racial minorities tend to assume operative-type positions in industry, then the supply of single white women might more than adequately meet the demand for clerical and service positions. In subsequent analyses, we plan to examine the possible influence these two variables have on the labor force participation of white women and to examine the labor force participation of black women.

Second, by aggregating male managers and professionals across types of industries, we could be overlooking possible variation in the demand for clerical help between industrial types. That the correlation between the proportion of managers and the proportion of clerical personnel is only 0.74 lends credence to this view. A more careful analysis of the occupational distributions within industries may reveal a clearer relationship between female employment and the male occupational structure.

REFERENCES

E. Almquist, "Women in the Labor Force," *Signs*, Summer 1977, *2*, 843–55.

M. T. Hannan and A. A. Young, "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series," in David R. Heise, ed., *Sociological Methods*, San Francisco 1977, 52–81.

M. Nerlove, "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, Mar. 1971, *39*, 359–96.

Valerie K. Oppenheimer, *The Female Labor Force in the United States*, Berkeley 1970.

U.S. Bureau of the Census, *1950, 1960, 1970, Census of Population*, Washington 1951, 1961, 1971.

U.S. Department of Labor, *1975 Handbook on Women Workers*, Washington 1975, 83–122.

DISCUSSION

MARYANN O'HAGAN KEATING, St. Mary's College (Notre Dame): My comments on Julie Matthaei's paper are more suggestive than critical. First, it must be emphasized that the historical development of the American two-career family is not a universal experience. Second, rather than suggesting that society constructs two different types of human beings, could not one say that the absolute advantage in bearing children and the relative advantage in homemaking have historically expressed themselves in certain forms of model household and employment agreements? Third, a critical thesis developed in Matthaei's paper is that the technology of commodity production caused women to enter the labor force to secure products for themselves and their families. She might explore how these developments altered the terms of trade within the family. Fourth, the paper suggests that women at present are locked into sex-typed jobs with work agreements that are obsolete with respect to their own identity and changed family relationships. As an overall conclusion, I must take a more pessimistic outlook in expecting the workplace to conform to the needs of the two-career family. The difficulties associated with such a family, given society's low premium on children, will necessarily be internalized, hopefully with both partners bearing the cost.

The major hypothesis of Daniel Quinlan's and Jean Shackelford's paper is that female employment is a doubly derived demand. It is first derived from the demand for males, who assume leadership and high-paying positions, the demand for which is derived from those secondary and tertiary industries experiencing the greatest increase in consumer demand. Because the greatest increase in female labor force participation since 1950 is in clerical work, it is hypothesized that female labor force participation in general can be explained by the percentage of males in managerial and professional work in the previous decade and by female participation in the secondary sector. The Quinlan-Shackelford empirical results failed to support this hypothesis, and they do not suggest an alternative hypothesis. Why? First, there is not enough variance in cross-sectional state data. Second, clerical workers may be experiencing a technological substitution of machinery and/or college-educated managers, who are themselves plentiful, relatively inexpensive, and willing to assume clerical and secretarial duties. Third, the Standard Industrial Classification of Occupations is much too aggregated to produce very meaningful results about manpower shifts. However, I am sympathetic to the basic idea of the paper. It is the case that in many fields such as medicine and education, much of the work is done by women in traditional paramedical and teaching occupations. However, these women earn one-third to one-half of their male supervisors who "check" to see if the teeth have really been cleaned, or the fractions taught. It is unfortunate that the Quinlan-Shackelford paper did not key in on those fields where quasi-monopolistic licensing, along with sex-typing, would confirm their hypothesis.

Investment and Growth in an Econometric Model of the United States

By ROBERT M. COEN AND BERT G. HICKMAN*

In its complete form, our Annual Growth Model is a synthesis of Keynesian and neoclassical determinants of actual and potential growth and of prices, production, and employment. In this paper we utilize only the supply side of the model to provide new quantitative estimates of the natural growth path of the *U.S.* economy since the Korean War and corresponding projections to the end of the century. The past and future contributions of business-fixed capital formation to productivity growth are quantified, and we assess the prospects for augmenting future growth through capital deepening. The complete model includes monetary and fiscal determinants of aggregate demand, but since we are abstracting from the demand side in this paper, nothing can be said here about nominal *GNP*, the price level, or the policy tradeoffs among inflation, unemployment, and growth targets.

I. Labor Supply and the Natural Unemployment Rate

The twin pillars of the present analysis are the labor supply and production sectors of the HC model. Both sectors have been expanded and reestimated since publication of the original model (1976).

Whereas we originally modeled labor force behavior entirely on the aggregate level, our new treatment disaggregates the labor force participation equations into sixteen age-sex groups. The participation rate for each of the eight age groups of males is a function of five aggregative variables: the ratios of civilian employment and of the armed forces to the noninstitutional population, the real after-tax wage rate, average annual hours per worker, and a time trend. One additional variable, the ratio of males aged 16–34 to those aged 35–64, appears in the female participation equations, in conformity with Richard Easterlin's hypothesis that changes in this ratio affect the participation rates of younger women positively, and those of older women negatively.

The real wage has generally insignificant effects on male participation rates, but the rates for young women are significantly negatively related to changes in the aggregate real wage. The principal effects of variations in aggregate employment and hours, included as indicators of cyclical changes in job opportunities and earnings at a given real wage, are on female participation rates. The participation rates for males 18–34 are negatively and significantly related to the size of the armed forces.

The equation for average hours is a hybrid of supply and demand factors. Workers' desired hours are a negative function of the real wage and the unemployment rate, with the latter proxying for cyclical variations in labor demand. Since part-time work is more prevalent among females, the ratio of women age 20 and over to the total labor force is included in the equation to capture the decline in average hours that occurs as the proportion of women increases.

The entire system of labor supply equations can be solved simultaneously for the "high-employment" supply of man-hours conditional on specified employment circumstances, the real wage, and population. We define *U7*, the high-employment unemployment rate, to be a weighted average

*Northwestern University and Stanford University, respectively. This research was supported by the National Science Foundation under Grant DAR77-27746.

of the unemployment rates by age and sex in a high-employment base year, 1956, the weights being the proportions of the high-employment labor force in each age-sex group. The rise in *U7* during 1953–77 reflects the growing mismatch between job requirements and job skills as the composition of the labor force increasingly shifted towards teenagers and women.

Our explanation of actual unemployment synthesizes the search and mismatch theories. Given the decision to participate in the labor force, one must still decide whether to retain one's job or to quit to search elsewhere, or if already unemployed, whether to accept new offers as they become available or to continue instead to look for a better opening. It is not implied that all unemployment represents a voluntary decision to await a better offer, however, since workers in depressed locations or who are lacking in skills or education may not receive offers despite lengthy searches. Our estimated equation for the actual unemployment rate, *U1*, discriminates between the transitory components stressed in search theory and the more enduring elements of the mismatch hypothesis.

In the estimating equation it is assumed that *U17*, the "natural" unemployment rate, which at a given time would be consistent with general equilibrium in the labor and commodity markets (see Milton Friedman), has been influenced by three principal factors over the postwar period. The first and dominant influence is the shift in the age-sex composition of the labor force, as measured in the high-employment unemployment rate *U7*. Subsidiary and marginally significant factors include variations in the relative size of the armed forces and in the ratio of average unemployment benefits to the private after-tax wage rate.

II. Production and Factor Demands

The basic theoretical structure of our system of factor demand and production functions (1970) remains unchanged. In the process of reestimating the system on data through 1978, however, we have introduced new formulations of output and price expectations, reexamined our earlier measures of labor input and of the rate of discount entering the implicit rental price of capital, and tested for recent decelerations in the rate of technical progress.

Assuming a Cobb-Douglas production function and cost-minimizing input decisions leads to a pair of equilibrium demand functions for labor and capital. Adjustment costs prevent immediate adaptation to changes in the equilibrium levels, however, and these costs are represented implicitly by partial adjustment hypotheses in logarithmic form. The resulting set of short-run *log*-linear relationships expresses the demands for capital and labor inputs as functions of their own lagged values and of expected output, the expected wage-rental ratio, and the trend rate of technical progress. The parameters include those of the underlying production function as well as the adjustment speeds for capital and labor, and the equations are jointly estimated by the maximum likelihood method subject to the parametric constraints across equations.

Just as there are short-term, disequilibrium demand functions for capital and labor corresponding to the long-run functions, so also is there a correspondence between the short- and long-term production functions. In the short-term production function the measured capital stock and measured labor inputs are corrected for cyclical changes in the degree of utilization or intensity of use of the measured inputs, whereas the measured inputs are evaluated at their normal or equilibrium intensities in the long-term version. A unique feature of our approach is to estimate the factor utilization intensities endogenously as part of the factor-demand system on the assumption that they are proportional to the degree of disequilibrium in each input (see our book, ch. 1, sec. 4).

In the empirical implementation, output is measured as gross private nonresidential product and the business fixed capital stock is measured net of depreciation at a constant exponential rate. Pollution abatement capital is included in the stock, since its exclusion had negligible effects on the parameter estimates. Labor input is mea-

sured in man-hours adjusted for changes in labor quality. The quality adjustment is similar to that used by George Perry, who weights each year's employment in each age-sex group by the product of average hours and earnings for the group in a base year, except that we need only adjust for the differences in average hourly earnings among the groups, since our man-hour measure of labor input already accounts for the differences in hours worked.

The wage-rental ratio appears in inverse form in the equations, with the nominal wage W measured by average hourly earnings before tax in the private sector. The implicit rental price of capital Q is a product of three terms: the price deflator for nonresidential fixed investment; the sum of the discount rate used in investment decisions and the depreciation rate; and a third term depending on the parameters of corporate taxation, including the investment tax credit and the tax treatment of depreciation allowances.

The variables entering the factor-demand equations are not actual but expected output and factor prices. The expectations of output, the nominal wage rate, and the investment deflator are determined by autoregressive equations estimated over the sample period. No allowance is made in Q for a difference between actual and expected discount rates or tax parameters.

After undertaking extensive tests on alternative measures of the discount rate, and on recent changes in the rate of technical progress, our preferred factor demand functions incorporate the following values for key parameters: the discount rate is a constant after-tax required rate of return of 10 percent; the long-run production elasticities of capital and labor are .31 and .69; the adjustment speeds of capital and labor are .17 and .71; and technical progress is estimated at a rate of 1.62 percent per year during 1950–68 and 1.03 percent during 1969–78. The hypothesis of an additional once-and-for-all downshift in the level of total factor productivity in 1974–75 is decisively rejected by the data. The hypothesis of an additional trend deceleration beginning in 1974 is also rejected at the 5 percent confidence level, with a t-ratio of only 1.5.

III. Characteristics of the Natural Growth Path

In the standard neoclassical growth model the natural rate of growth is determined by the demographically fixed growth rate of the labor force and the rate of Harrod-neutral technical progress. The saving rate determines the supply of capital and the wage-rental ratio adjusts to clear the labor and capital markets and assure full employment of both factors. An analogous approach could be taken in simulations of our complete model, which generates prices and saving endogenously, but for present purposes only potential supply is considered. The problem is to generate internally consistent relationships among the principal endogenous variables when looking at the supply side alone.

Assume initially that the full-employment path of man-hours is given. The labor demand function relates man-hours demanded to output, the wage-rental ratio, the technical progress trend, and lagged man-hours. In our model, "full-employment output" is determined by solving the labor demand function for output and substituting full-employment man-hours for actual man-hours (see 1976, ch. 1, sec. 5). This expression may be converted to an equation relating full-employment man-hour productivity to the wage-rental ratio, the rate of change of full-employment man-hours, and technical progress. Since technical progress is exogenous and the growth rate of man-hours is assumed given, the only unknown needed to determine man-hour productivity, and hence output itself, is the wage-rental ratio. In order to avoid exogenous extrapolation of this key variable, we assume its real-wage component, the ratio of the investment price deflator and the nominal wage rate, to be endogenous. We impose the condition, consistent with our Cobb-Douglas specification and the price equation in the complete model, that the growth rate of the real wage equals that of man-hour productivity. Since the real wage also affects labor force par-

TABLE 1—NATURAL UNEMPLOYMENT AND GROWTH RATES AND INVESTMENT RATIOS, SELECTED PERIODS, 1955–2000
(Shown in Percent)

Period	Natural Unemployment Rate[a]	Natural Growth Rates[b]: Output	Man-Hours	Man-Hour Productivity	Output per Capita	Capital Stock	Investment Ratios[a]: Net	Gross
1955–68	4.2	3.5	0.7	2.8	2.0	4.3	2.3	9.2
1968–73	4.6	3.2	1.8	1.5	2.3	3.9	2.3	9.8
1973–78	4.9	2.9	1.9	1.0	2.2	3.0	1.8	9.5
1978–83	4.8	3.3	1.6	1.7	2.4	3.2	1.9	9.5
1983–92	4.6	2.7	0.9	1.8	1.8	2.9	1.8	9.5
1992–2000	4.4	2.5	0.9	1.5	1.8	2.6	1.6	9.5

[a]Averages for period. Investment ratios are shares of full-employment *GNP*.
[b]Exponential rates between endpoints of periods.

ticipation and hours of work, this assumption also makes the relative price component of full-employment labor supply endogenous. Finally, the growth path of capital stock is fully determinate, since it depends on full-employment output, the wage-rental ratio, and technical progress.

We have solved the entire system simultaneously for 1953–2000, using as initial conditions the actual 1952 values of man-hours and capital stock. The transitory effects of using actual instead of full-employment values as initial conditions are quickly absorbed and the natural growth path is firmly established by 1955. Actual values of the exogenous variables are used through 1977 or 1978 and later observations are extrapolations or projections of various sorts. The population projections are from the Bureau of the Census. The exogenous components of the wage-rental ratio and the armed forces and unemployment insurance variables are held constant after 1977 in the basic simulation.

One further adjustment was necessary to account for the exogenous upshift in the rental-wage ratio in 1973–75 as a result of the quadrupling of oil prices by *OPEC*. This external shock raised the ratio of the investment price deflator to the wage rate by 6.5 percent between 1973 and 1975, after which time the endogenous downtrend resulting from productivity growth was resumed, and we have introduced a corresponding once-and-for-all upshift in the path of the rental-wage ratio in those years.

The principal results are summarized in Table 1. The natural unemployment rate rose from 3.9 percent in 1955 to 5.0 in 1975 and averaged 4.9 during 1973–78. It is projected to decline gradually through the end of the century.

The natural growth rate of output decreased from 3.5 percent in 1955–68 to 3.2 in 1968–73 and 2.9 in 1973–78. Productivity growth fell markedly in 1968–73, but the impact on output growth was mitigated by a sharp acceleration of man-hour growth. Man-hours again accelerated slightly in 1973–78, but a continuing decline in productivity growth further depressed the growth rate of output in those years.

Without the oil shock, the growth rate of potential productivity would have been 1.7 instead of 1.0 percent per annum in 1973–78. The potential growth path would have been 3.0 percent higher by 1978, but the subsequent natural growth rates would have been unchanged from those in Table 1. Together with the fact that neither an additional trend break nor a once-and-for-all downshift in total factor productivity in the mid-1970's proved to be statistically significant in our factor-demand functions, this result implies that the principal effects of the energy shock on potential output growth

TABLE 2—CONTRIBUTING SOURCES AND ANNUAL RATES OF CHANGE OF POTENTIAL MAN-HOUR PRODUCTIVITY, HISTORICAL 1955–78 AND PROJECTED 1979–2000 (Shown in Percent)

Period	Productivity	Technical Progress	Labor Quality	Labor Utilization	Capital Utilization	Capital Deepening
1955–68	2.82	1.62	0.01	0.00	0.10	1.09
1968–73	1.46	1.03	−0.19	0.04	−0.11	0.69
1973–78	1.04	1.03	0.00	0.01	−0.41	0.40
1978–83	1.73	1.03	0.23	0.04	−0.04	0.48
1983–92	1.81	1.03	0.24	−0.02	−0.10	0.65
1992–2000	1.53	1.03	0.00	0.02	−0.03	0.51

were transitory and attributable to the resulting rise in the relative price of capital.

Our projections indicate that potential productivity growth will accelerate in 1978–83 and will increase the natural growth rate to 3.3 percent despite a deceleration in man-hour growth. After 1983 the natural rate is projected to resume its decline, largely because man-hour growth will slow.

Net investment requirements along the natural growth path fell in 1973–78 as growth decelerated. A similar reduction appears in the gross investment share. The net share is projected to decrease somewhat in the 1980's and 1990's, but the gross share will remain constant as replacement investment rises moderately along with the capital-output ratio. Thus it appears that nonresidential fixed investment requirements will not in future constitute an increasing claim on national resources at full employment as they did prior to 1973.

The contributions of capital deepening and other factors to the natural rate of productivity growth are shown in Table 2. Deepening was second only to technical progress as a source of potential productivity growth during the postwar years and is projected to remain so along the natural path. The major sources of the reduction of potential productivity growth during 1968–73 were the reductions in the rates of technical progress and of capital deepening, with secondary contributions from decreases in labor quality and the degree of normal capacity utilization. The further reduction of potential productivity growth after 1973 reflects the decline in the natural rates of capital deepening and capital utilization resulting from the *OPEC* shock to the rental-wage rate and the accompanying deceleration of the natural growth rate. Conversely, potential productivity growth is projected to increase as these negative factors are reversed in coming years.

What is the scope for policy actions to increase the rate of capital deepening to raise productivity growth? The secular downtrend in the rental-wage ratio is fundamentally governed by productivity growth itself, but the exogenous component of the ratio can be altered by fiscal policy although not, in our formulation with a constant required rate of return of 10 percent, by monetary policy. If thought to be permanent, a once-and-for-all reduction in the rental-wage ratio will induce a transitory increase of productivity growth as the economy moves to a corresponding higher natural path. To judge the quantitative importance of this policy option, we have introduced a permanent 20 percent reduction in the rental price of capital beginning in 1980, equivalent to more than halving the corporate tax rate, and recalculated the natural growth path through 2000. The transitory effects on productivity growth are largely absorbed within three years. Along the original path productivity increases at a rate of 1.8 percent between 1979 and 1983, whereas the rate would be 3.9 percent under the new policy. The post-1983 natural growth rates would be virtually the same on either path, but the policy change would have raised the level of the natural path

permanently by 10 percent, at the cost of a permanent increase in the gross investment ratio from 9.5 to 11.8 percent. The natural rate of unemployment would be permanently reduced by .05 of a percentage point.

IV. Conclusion

We have presented new estimates of the natural growth path of the *U.S.* economy using a disaggregated model of labor supply and unemployment and an integrated factor-demand and production system. To our knowledge, this is the first such application in which the natural growth paths of output, capital stock, man-hours, labor productivity, and the real wage and rental-wage ratios are all endogenously determined and internally consistent.

Rates of capital formation higher or lower than the natural rates indicated by the solution of our system would conflict with the maintenance of labor market equilibrium at the natural rate of unemployment. A permanent reduction in the level of the rental-wage ratio, brought about for example by a cut in the tax rate on profits, can increase the natural rates of capital formation and output growth, but for only a few years. The system soon settles down to the same growth rates, albeit with higher output levels. The short-lived increase in the natural growth of output entails, however, a permanently higher ratio of capital to output and of gross investment to *GNP* and a slightly lower natural unemployment rate.

The specific numerical results obviously depend on the estimated parameters. If subsequent observations confirm that a further deceleration or downshift of technical progress occurred in the mid-1970's—a hypothesis so far rejected by the data at the usual significance levels—our estimated growth rates of labor productivity and output after 1973 would be lower by about one-half a percentage point. Recall also that the estimates refer to movements along the natural growth path. Although on the path in 1973, the economy has dropped well below it in subsequent years, with consequent short-falls in the accumulation and utilization of capital and in realized productivity growth. Fortunately these shortfalls can eventually be eliminated, but about four or five years of vigorous recovery and investment would be required to regain the natural path.

Finally, although we have referred to our solution as a natural growth path, it is not the steady-state concept of conventional growth theory, which requires the condition that the real wage per man-hour increase at the rate of Harrod-neutral technical progress instead of the rate of man-hour productivity. It is easy to obtain such a steady-state solution to our model, but the concept is unnecessarily restrictive and less meaningful as a benchmark for potential output than the one we have adopted.

REFERENCES

R. M. Coen and B. G. Hickman, "Constrained Joint Estimation of Factor Demand and Production Functions," *Rev. Econ. Statist.*, Aug. 1970, *52*, 287–300.

R. A. Easterlin, "What Will 1984 Be Like? Socioeconomic Implications of Recent Twists in Age Structure," *Demography*, No. 4, 1978, *15*, 397–432.

M. Friedman, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

Bert G. Hickman and Robert M. Coen, *An Annual Growth Model of the U.S. Economy*, Amsterdam 1976.

G. L. Perry, "Potential Output and Productivity," *Brookings Papers*, Washington 1977, *1*, 11–47.

Bureau of the Census, *Current Population Reports*, Series P-25, *Projections of the Population of the United States: 1977 to 2000*, No. 704, Washington 1977.

The Contribution of Capital to Economic Growth

By EDWARD F. DENISON*

How important to economic growth in advanced countries is the accumulation of physical capital? My short answer is that increased capital is *one* of *several* important sources of output growth. This appraisal, which is amply supported by research results, should surprise no one—but doubtless it will.

Capital is not *the* source of growth despite the contrary view common in financial circles and on Capitol Hill. Output, and therefore growth, are governed by *many* determinants. Moreover, there is no substance to the recurring notion that the United States' growth rate would have matched those of countries like Germany or Japan if only we invested as much as they did. Finally, it is quite wrong to blame investment for the recent sharp reduction in the growth of American output and productivity, or to suppose that merely raising investment would go far toward restoring the old growth rate of productivity.

Why many people share a vision of growth that assigns exclusive attention to capital I do not know. Growth models that feature capital, while assigning other output determinants to *ceteris paribus*, may be partly responsible, but it should have been apparent to all that such models were meant only to illuminate a relationship, not to describe a whole economic system. Analyses based on incremental capital-output ratios may have contributed to the illusion. If so, as Robert Solow says of these ratios, "Economists have a responsibility to do better."

To deny that capital is everything is not to imply that it is nothing. I do not share the other extreme view, sometimes encountered, that capital can be ignored because its significance is hard to establish if one fits a production function by correlation analysis. I stress again: capital is an important growth source. It has sometimes contributed importantly to differences in growth rates between periods and places. More capital formation would raise the growth rate.

The contribution of capital to growth is evaluated best in the context of a complete analysis of the sources of growth. The following summary, drawn from my *Accounting for Slower Economic Growth*, refers to total potential national income originating in nonresidential business. During the period from 1948 to 1973 the growth rate of this series was 3.8 percent per year. Of that amount, 15 percent resulted from more capital, that is, more nonresidential structures and equipment and more inventories. Another 15 percent is ascribed to changes in employment and working hours, with account also taken of the age-sex composition of workers. Fourteen percent was due to increased capabilities of workers resulting from more education. Ten percent resulted from improved resource allocation, taking the form of a reduction in the amount of labor overallocated to farming and to self-employment and unpaid family labor in nonfarm establishments too small for efficiency. Thirty-seven percent was contributed by advances in technological, managerial, and organizational knowledge as to how to produce at low cost, together with miscellaneous output determinants not separately estimated. This is the residual in the calculation. In the 1948–73 period it probably provides a tolerable approximation to the contribution of advances in knowledge alone. If so, advances in knowledge were much the largest single source of growth. Economies of scale made possible by the growth of markets contributed an estimated 11 percent of the growth rate. Finally, certain changes in the legal and human environment, together with irregular factors, subtracted 2 percent.

*Associate Director for National Economic Accounts, Bureau of Economic Analysis.

These estimates refer to potential national income in the nonresidential business sector, which omits nearly one-fourth of national income. To consider the whole economy we must bring in the net output of dwellings and net property income received from abroad, both entirely ascribable to capital, and the output of workers employed in general government, households, and institutions, none of which is ascribable to capital given present methods of output measurement. Capital was the source of 18 percent of the growth rate of potential national income in the economy as a whole as against 15 percent in nonresidential business.

With rising employment, capital input per worker increases less rapidly than total capital input. As a result, capital contributes fewer percentage points to growth of output per worker than of total output. From 1948 to 1973, capital contributed an estimated 0.43 percentage points to the growth rate of potential national income per person potentially employed as compared with 0.71 percentage points to the growth rate of total potential national income. However, growth rates of output are also lower on a per-person basis. It turns out that in nonresidential business, capital's percentage contribution was the same, 15 percent of the growth rate, on a per-person basis as in the aggregate, while in the whole economy it was 19 percent as against 18. All four percentages are raised, but only a point or two, when calculations are on an actual instead of a potential output basis. Among the eight series I have considered, capital's contribution to growth ranged from 14 to 21 percent in 1948–73. Use of an alternative classification that reallocates gains from economies of scale among the other growth sources instead of considering them a separate growth source would raise the percentages for the contribution of capital, those for nonresidential business by as much as one-eighth. They would then range from 16 to 23 percent in 1948–73. On the other hand, starting with 1929 instead of 1948 lowers all the capital percentages appreciably.

In summary, capital has been responsible for one-seventh to three-thirteenths of the long-term growth of net output in the United States, the exact amount depending on the time period analyzed, the classification of economies of scale, and whether one is considering total or per-worker output, potential or actual output, and the whole economy or nonresidential business. My methods of estimation have been fully described elsewhere. They rest mainly on the reasonable presumption that enterprises try to combine factors of production in such proportions as to minimize costs.

The impression that the United States needed only to match the investment ratios of Germany or Japan in order to match their growth rates goes back at least to a widely publicized 1960 report by Nelson Rockefeller. Much later William Simon, as Secretary of the Treasury, fostered the same idea and it has remained prevalent until now. Rockefeller's and Simon's evidence was correlation between growth rates and investment ratios observed in a comparison of a small number of countries. The idea seems impervious to repeated showings that the correlation dwindles away if the exact selection of countries or the exact definition of investment is changed; that the causal inference drawn from the correlation lacks a theoretical basis since growth models show growth rates to be independent of levels of investment ratios; and that detailed research has shown that international differences in growth rates are due mainly to output determinants other than capital. Moreover, by examining chronological sequences, Richard Franke has recently confirmed my long-standing belief that whatever relationship does exist results more from the effect of rapid growth on investment than from the effect of investment on growth.

The exaggeration in the idea that investment ratios can account for international differences in growth rates is obvious if one takes the trouble to ponder a few numbers. In the postwar years up to 1973, net private investment averaged about 7.5 percent of net output. Suppose it had been higher. At, say, 1969 output levels, each additional 1 percent of the national income invested would have provided \$7.7 billion of additional capital. Net earnings of capital, before tax, equaled 8.0 percent of net asset

value. If the rate of return had been this high for the additional capital, then each one-point increase in the investment ratio would have raised the annual national income by 8.0 percent of $7.7 billion. This comes to $0.6 billion or 0.08 percent of the national income. If the extra capital were all allocated to nonresidential business, where the ratio of earnings to assets was highest, 10.4 percent, the national income would have been raised more—by 0.11 percent. Let us not only assume this more favorable allocation, but also raise the 0.11 percent to 0.12 to allow for economies of scale. Under these conditions, an increase in the net investment ratio of 1 percentage point would raise the growth rate of national income by 0.12 percentage points. An increase of 8.5 percentage points in the net investment ratio would then have been required over an extended period to add 1 full percentage point to the growth rate over that period. This would have more than doubled the actual net investment ratio. To have added 6 percentage points to the growth rate, as would have been required to match Japan, the net investment ratio would therefore have had to be raised six times this much, more than 50 percentage points, which would have brought it to 60 percent of national income.

Actually, very high investment ratios would aid long-term growth rates much less than these calculations suggest. A large addition of capital, in the absence of similar additions to labor and land, would reduce the relative marginal product of a unit of capital and lower the rate of return. Other qualifications and adjustments are needed but, in general, refinements only intensify the thrust of the results. In practice the United States could have eliminated only a small part of growth rate differentials by investing more.

This conclusion is supported by comprehensive studies of the sources of growth in eleven countries, covering various time periods. Five of the other ten countries, all relatively large economies, had growth rates well above the United States. In two of the five, Italy and France, capital contributed no more to growth than in the United States so it explains none of the difference in growth. In three of the five, capital contributed more to growth than in the United States but the difference in the capital contribution was just over one-fourth as large as that in growth rates in the cases of both Japan and West Germany, and just over one-third in the case of Canada. A sixth large country, the United Kingdom, had a lower growth rate and a smaller capital contribution than the United States, with the difference in the capital contribution less than one-fifth as large as the difference in growth rates. Comparisons with the four smaller countries show no particular pattern. As these illustrations show, capital falls far short indeed of fully explaining international differences in growth rates, although it is often a factor. The sources of remaining differences in growth rates have been detailed elsewhere. Japan, which had by far the highest growth rate, secured much larger contributions not only from capital but also from changes in average working hours, the age-sex composition of employment, and resource allocation, from the incorporation of new knowledge into production, and from economies of scale.

In the United States itself, output has grown much less rapidly since 1973 than it had from 1948 to 1973. I have calculated the contribution of capital to growth in the period 1973–78 in order to ascertain how much of the decline in output growth is ascribable to capital. The results do not show capital contributing a disproportionately large amount to the decline in growth; indeed, its importance in the decline was less than its importance in 1948–73 growth.

Consider first the drop in the growth rate of total actual national income in the whole economy, which was from 3.65 percent in 1948–73 to 2.38 percent in 1973–78. Among the four types of capital that affect this series, retardation was greatest for inventories. The growth rate of the stock of inventories fell from 3.48 percent to 2.23 percent. The contribution of inventories, which is the product of their growth rate and their weight, fell even more sharply because the share of inventories in total earnings in the economy as a whole, which establishes their weight, also declined even after elimination

of cyclical influences. However, even before the decline inventories carried a weight of only about 3 percent of total input and contributed only 0.12 percentage points to the growth rate. The drop was to 0.05 points in 1973–78, so 0.07 points of the 1.27 percentage point decline in the growth rate of national income can be assigned to inventories. The growth rate of capital input in the form of nonresidential structures and equipment fell by only one-eighth, from 3.70 percent in 1948–73 to 3.24 percent in 1973–78. It should be observed that only in comparison with the late 1960's and early 1970's is the 1973–78 rate low; it is about the same as the rate from 1948 or 1950 to 1964. The weight of nonresidential structures and equipment also declined but, even so, their contribution fell only 0.08 points, from 0.32 to 0.24. The contributions of the two remaining types of capital can be obtained, with only minor adjustments, directly from the national income and product tables of the Bureau of Economic Analysis. Dwellings, which are the most important type of capital if judged by value, contributed *more* to growth in 1973–78 than in 1948–73, 0.27 percentage points as against 0.24. The contribution of the final capital component, international assets, fell from 0.03 percentage points to 0.01. The total contribution of the four types of capital dipped from 0.71 percentage points in 1948–73 to 0.57 in 1973–78. The drop, 0.14 percentage points, equals 11 percent of the 1.27 percentage point drop in the growth rate of national income. Capital had been the source of 19 percent of the 1948–73 growth rate of this series so it contributed less than proportionately to the retardation.

This is also the case if potential national income in the whole economy, or actual or potential national income in the nonresidential business sector, are considered. Of the declines in the growth rates of these series, 9 to 14 percent is attributable to capital.

A feature of the post–1973 years was fast employment growth. Actual employment increased 2.0 percent a year from 1973 to 1978 as against 1.5 percent in 1948–73. Consequently, capital input per person employed slowed more than total capital input. This meant that, when measured in percentage points, capital's contribution to the growth rate of national income per person employed fell more than its contribution to the growth of aggregate national income: by 0.19 points as against 0.14. But the growth rate of national income per person employed also fell more, from 2.1 percent in 1948–73 to only 0.39 percent in 1973–78. Thus capital was responsible for 11 percent of the decline, the same percentage as for aggregate national income. In contrast, capital had contributed 21 percent of the 1948–73 growth of this series. Growth of potential employment accelerated even more than actual employment, from 1.6 percent a year in 1948–73 to almost 2.4 percent in 1973–78, and among all the series explored the drop in capital's contribution was biggest, 0.35 percentage points, for potential national income per person potentially employed in the whole economy. This still was only about 17 percent of the drop in that growth rate, less than capital's contribution to growth of the same series in 1948–73.

A predictable reaction to such results is: "Yes, but they do not take into account changes in the age distribution of structures and equipment." Businessmen sometimes imagine the supposed omission to be very important, but it is not. The crucial assumption of the extreme vintage models from which that belief derives is that reducing the average age of structures and equipment by one year means picking up one year's technical progress. Even if this were correct, changes in average age could account for no more than two-tenths of a percentage point of the growth slowdown in nonresidential business and less in the whole economy. But as Arnold Harberger and I independently showed back in 1964, the assumption is inadmissible. Amounts of quality improvement vary greatly among types of capital goods. The return on replacement investment is highest, and hence the incentive to invest greatest, for types that have suffered the most obsolescence as a result of quality improvement in newer vintages. Any substantial amount of gross investment permits the investment opportunities created by sizable quality improvements to be grasped.

Additional investment is less-profitable investment devoted to types of capital goods in which quality change has been small. The gain that vintage models imagine to be derived from additional new investment is not realized because a reduction in the average amount of quality improvement incorporated in new capital automatically offsets the lowering of the average age.

There are various ways in which the importance of capital in growth can be made to appear larger or smaller. Focusing on gross rather than net product, for example, automatically raises the percentage of growth ascribed to capital because capital consumption is counted as a contribution of capital. On the other hand, the Cambridge concept of capital as consumption foregone lowers the capital contribution; in *Accounting for Growth* I estimated that it transferred 0.2 percentage points from capital to advances in knowledge in 1948–69. Alternatively the state of knowledge, education, and other social characteristics can be described as human or social capital. But none of this is substantively important.

The growth of total output has declined by a large amount since 1973, and growth of output per person has almost stopped. National income per person employed in 1979 was barely above 1973 on either an actual or a potential basis. The adverse trend is important, primarily because of its effect on living standards. It also contributes to other problems, particularly inflation. By its effect on inflation, it makes it harder to deal with unemployment. Shrinking the pie available for division among people and programs, it intensifies conflicts over the economic issues that divide the country.

The causes of the productivity setback are numerous and only partly understood. I suspect the slowdown will prove stubborn and difficult to reverse. Under these conditions it is frivolous to teach that a moderate increase in investment, induced by some painless incentives to invest or save more, would restore past growth rates, or that raising investment ratios to those in Germany or Japan would give us growth rates like those countries enjoyed before 1973. It is not only frivolous, but also a dangerous diversion. We shall not escape our predicament without a broad program that affects numerous output determinants and requires actions less pleasant than cutting taxes on investment, saving, or income from capital. Again, I must stress that I favor trying to raise investment, although I am skeptical that we know how to obtain a very big increase. But raising investment may also be the least fundamental part of the problem of growth and productivity because the observed setback to capital stock growth after 1973 was mainly a reaction to unemployment and unutilized capacity. It was, that is to say, more the consequence of other problems than their cause.

REFERENCES

Edward F. Denison, *Accounting for Slower Economic Growth: The United States in the 1970s*, Washington 1979.

______, "The Contribution of Capital to the Postwar Growth of Industrial Countries," in U.S. Congress, Joint Economic Committee, *U.S. Economic Growth from 1976 to 1986: Prospects, Problems, and Patterns*, Vol. 3, *Capital*, pp. 45–83.

______, "The Unimportance of the Embodied Question," *Amer. Econ. Rev.*, Mar. 1964, *54*, 90–94.

R. H. Franke, "Investment for Economic Growth?," paper presented to American Association for Advancement of Science, Houston, Texas, Jan. 3, 1979.

A. C. Harberger, "Taxation, Resource Allocation, and Welfare," in *The Role of Direct and Indirect Taxes in the Federal Revenue System*, Princeton 1964.

N. A. Rockefeller, "Accelerated Economic Growth: A Key to the American Future," in *Governor's Conference on Automation*, New York, June 1960.

Robert M. Solow, *Capital Theory and the Rate of Return*, Amsterdam 1963.

N.Y. Stock Exchange Office of Economic Research, *Reaching a Higher Standard of Living*, Jan. 1979.

Total Income, Total Investment, and Growth

By ROBERT EISNER*

Productivity is declining, growth is stunted, and capital formation is inadequate. This, at least, is fast becoming the byword of the Halls of Congress, the business press, and, perhaps, the economics profession. But what do these words mean? At what levels, if any, are the assertions true?

What happens to income, output, and productivity when clothes washing moves from the wash tub and the professional laundry to the laundromat and to the automatic washer and dryer in the home? What happens when women respond to the availability of pre-prepared foods and labor saving devices in the household to move into the labor force and employment of less than average market productivity? Are we more productive if we devote proportionately less of our resources to national defense or more to police services? What happens to productivity, income, and output if we watch films on our home television screen rather than in the local movie house?

Or is capital formation less when we acquire more of our own private automobiles as orders for new busses decline? Is capital formation higher if we rent more cars from Hertz and Avis? Is capital formation more when we spend more on new plants and equipment but less on research and development or education? Is the "productivity" of capital reduced if larger proportions of it are directed to giving us cleaner air and water and less to market output?

In sum, are there dimensions to income, output, productivity, and investment so broadly ignored that judgments about productivity decline or inadequate capital accumulation (let alone their causes) are of limited relevance to the analysis of economic well-being and progress?

Following long-established concepts and insights of Simon Kuznets, I have, with a number of research associates, been developing and filling in "*TISA*," a total incomes system of accounts. This project profits from the work of James Tobin and William Nordhaus, John Kendrick, Richard Ruggles and Nancy Ruggles, F. Thomas Juster, and Oli Havrylyshyn, among others. Contrasting with the traditional Bureau of Economic Analysis (*BEA*) framework, we endeavor to include all economic output, whether market or nonmarket, and all capital accumulation, whether by government, households, or business, and whether tangible or intangible. In so doing, we take final output to include all consumption services, those produced in the household, furnished by government, or purchased in the market, domestic capital accumulation of all sectors, and net exports. Government production is thus allocated to consumption, investment, and intermediate product as reported in my earlier work (1978) and my paper with David Nebhut. The preparation of *TISA* leans heavily on the magnificent income and product accounts and data of the *BEA*.

Abandonment of the concept of final purchase as a definition of final product leads us to include uncompensated factor services of volunteers, military draftees, and jurors, and to reclassify media services of commercial television, radio, and newspapers as consumption rather than intermediate product. Moving in the opposite direction, the services of police and national defense are reclassified as intermediate.

We have worked with masses of *BEA* data, unpublished as well as published,

*Professor of economics, Northwestern University. The extensive work underlying this paper has enjoyed continuing financial support of the National Science Foundation and generous repeated assistance of members of the staff of the Bureau of Economic Analysis. Research associates most actively involved in development and processing of the data for the current version of our accounts are David H. Nebhut, Emily R. Simons, Marsha J. Courchane, Paul J. Pieper, Stuart E. Weiner, David A. Reishus, and Bruce D. Meyer.

along with data from other sources. We have undoubtedly strained credulity with many heroic assumptions, and with interpolations and extrapolations for missing data points in a number of series. But essentially we have endeavored to be consistent with the economic theory of income, output, consumption, and capital accumulation. We have measured consumption as household production or receipt of consumer services and nondurable goods. We have measured tangible capital formation as the cost of acquisition or production of fixed capital in all sectors, including consumer durables acquired by households and some semidurables such as clothing.

Since our ultimate concern has not been the market price of market output, we also include estimated capital gains and losses on existing tangible assets, that is, revaluations net of general changes in the price level, as reported in my 1980 paper. These are available for inclusion in measures of total income as well as total capital accumulation. Similarly, subsidies and deficits of government enterprises are added to the value of product, included partly in investment and partly in consumption.

Work of Theodore Schultz, Gary Becker, Edward Denison, Fritz Machlup, and many others has demonstrated the importance to output and productivity of the vast amounts of economic activity directed to what may be called intangible capital formation. We have prepared estimates based on both expenditures and the opportunity costs of time and resources devoted to investment in research and development, education and training, and health.

My earlier work reported total incomes current dollar accounts for 1959 and 1969 (1978), capital gains and losses for all of the years from 1964 to 1977 (1980), and current and constant dollar estimates for the government sector for 1946 to 1976 (my paper with Nebhut). In ongoing work we are now completing estimates of current and constant dollar accounts for all of the sectors for all of the years from 1946 to 1976. Some of the findings appear instructive and, despite their highly preliminary character, with various revisions and discrepancies still to be taken care of, I should like to share them with you.

First, it becomes apparent that the usual measures of business investment vastly understate more broadly defined capital formation. Gross *private* domestic investment reported by the Bureau of Economic Analysis comprises business expenditures for new plant and equipment, including rental housing, expenditures by households for new owner-occupied housing, nonprofit institution investment, business acquisition of used capital, and change in business inventories. The *BEA* figure for 1976 was \$243 billion. By contrast, my estimate with Nebhut of total capital formation related to government, exclusive of net revaluations and the value of military and other personnel training, came to \$248.6 billion, some 2.3 percent more.

Not all of the capital formation related to government, to be sure, represents tangible capital. Government and government enterprise fixed investment in 1976 was \$60 billion, and the change in government inventories was \$5 billion. In addition, we estimated that the government sector accumulated \$23 billion of product in the form of investment in research and development and natural resources. Government also gave some \$14 billion of capital to business enterprises, essentially research and development output. And most important, government production of capital assumed distributed to households in the form of education and training came to \$128 billion. Addition of half of government health product, taken as capital formation, brings \$18 billion to the total.

Viewed in constant dollars, the government output going to capital formation has been growing throughout the post-World War II period, but at a decreasing rate since the 1950's. Taking average percentage growth rates over successive five-year periods, the progression, beginning with 1951–55 to 1956–60 and ending with 1966–70 to 1971–76, was 9.1 percent, 6.1 percent, 5.8 percent, and 3.8 percent. Amidst widely expressed concern for allegedly lag-

TABLE 1—GROSS CAPITAL ACCUMULATION, PRELIMINARY ESTIMATES, 1976

	Billions of Dollars	Percent of Original Cost Total
Gross Domestic Capital Accumulation	1,301.9	116.6
Original Cost Total	1,116.1	100.0
Tangible	558.3	50.0
Gross Private Domestic Investment	243.0	21.8
Fixed	225.1	20.2
Business	167.0	14.9
Nonresidential	156.6	14.0
Residential	10.4	0.9
Owner-occupied residential	52.2	4.7
Nonprofit institutions	5.9	0.5
Change in inventories	10.2	0.9
Plus discrepancy[a]	7.7	0.7
Other Tangible Investment	315.3	28.2
Fixed	309.5	27.7
Government (including gov't. enterprises)	75.8	6.8
Households	241.4	21.6
Less discrepancy[a]	7.7	0.7
Change in inventories	5.8	0.5
Intangible	557.8	50.0
Research and development	37.7	3.4
Education	444.9	39.9
Health	75.1	6.7
Subsidies Allocated to Investment	3.7	0.3
Net Revaluations	182.0	16.3
Fixed reproducible capital	57.7	5.2
Inventories	−3.4	−0.3
Land	127.7	11.4

[a]Due essentially to valuation of net purchases of used goods at market prices in income and product accounts, and at undepreciated historical costs in investment by legal type of ownership in the *BEA* capital stock series.

ging rates of growth of business investment, one might note the distinct lag in growth rates of government output going to capital.

Capital formation by households is also very large relative to that by business. Household investment in durable goods alone amounted to $157 billion in 1976. Investment in semidurables came to another $85 billion while investment in owner-occupied nonfarm dwellings, included in the *BEA*'s residential construction and gross private domestic investment, was $52 billion. Business fixed investment, excluding owner-occupied nonfarm dwellings and investment by nonprofit institutions, was only some 30 percent of total tangible investment.

Inclusion of intangible capital accumulation reduces the business fixed investment share of the total all the more. We estimate $445 billion invested in education in 1976, of which $203 billion were opportunity costs of students. Intangible investment in research and development came to $38 billion. If we assume that half of health services constituted investment in health capital, we add another $75 billion to intangible capital accumulation.

Net revaluations totalled $182 billion, of which $128 billion were on land and $54 billion on reproducible capital. Even leaving out net revaluations, it can be seen in Table 1 that the *BEA* gross private domestic investment of $243 billion was less than 22 percent of gross domestic capital accumulation. Business nonresidential investment in structures and equipment, upon which some would apparently wish to lavish so much in

TABLE 2—SELECTED COMPARISONS OF *BEA* AND PRELIMINARY *TISA* ESTIMATES, 1976, AND RATES OF GROWTH, 1946–76[a]

	Billions of Dollars	Percent Rates of Growth			
	1976	1946–56	1956–66	1966–76	1946–76
BEA Gross National Product	1,700.1	7.2	6.0	8.5	7.2
TISA Gross National Product	2,937.4	8.3	5.2	9.9	7.8
BEA as Percent of *TISA*	57.9				
BEA Net National Product	1,522.3	6.9	6.1	8.2	7.1
TISA Net National Product	2,280.8	9.4	5.2	9.8	8.1
BEA as Percent of *TISA*	66.7				
BEA National Income	1,359.2	6.9	6.0	8.1	7.0
TISA Net National Income Minus Net Revaluations	2,062.2	6.1	5.6	8.7	6.8
BEA as Percent of *TISA*	65.9				
BEA Net Private Domestic Investment	65.2	6.7	6.9	0.4	4.6
TISA Net Domestic Capital Formation (at Original Cost)	459.5	31.1	8.0	7.8	15.1
BEA as Percent of *TISA*	13.3				
TISA Net Tangible Capital Formation	129.0	b	8.0	1.2	b
TISA Net Intangible Capital Formation	330.4	10.1	8.1	10.9	10.6
TISA Gross Domestic Capital Formation	1,301.9	15.7	5.3	10.9	10.6
BEA Gross Private Domestic Investment	243.0	8.8	5.8	6.9	7.1
TISA Capital Formation Related to Government, Exclusive of Net Revaluations	248.6	b	6.6	9.3	b
TISA Government Capital Formation as Percent of *BEA*	102.3				
BEA Personal Consumption Expenditures	1,090.2	6.3	5.7	8.9	7.0
TISA Consumption	1,628.1	5.6	5.1	9.1	6.6
BEA as Percent of *TISA*	67.0				

[a]*BEA* figures are first revisions from July 1978 *Survey of Current Business* to preserve consistency with *TISA* components.
[b]Dollar figure negative in 1946.

new tax advantages, was then only some 14 percent of that total, \$157 billion out of \$1,116 billion.

Comparisons of measures of gross capital accumulation can be particularly misleading where length of life of assets and depreciation rates differ. Capital consumption allowances on investment in household semidurables and short-lived durable goods will be a high proportion of gross investment. The difference between new capital accumulation and the narrow measure of net investment usually considered, however, is only moderately less than the difference in gross accumulation figures. As indicated in Table 2, total net tangible capital formation, excluding net revaluations, came to \$129 billion for 1976 as against net private domestic investment of \$65 billion.

Turning to measures of income and product, we find the 1976 *BEA GNP* of \$1,700 billion only 58 percent of our \$2,937 billion preliminary estimate of *TISA GNP*. The latter comprises \$1,628 billion of consumption, \$1,301 billion of gross domestic capital accumulation and \$7 billion of net exports. Swelling our estimates are \$342 billion of consumption services of durable and semidurable goods owned by households, \$414 billion for unpaid household labor services,

$203 billion for opportunity costs of students (imputed to capital formation), $21 billion for business transfer payments in the form of media support, and $3 billion in safety and health services provided by business. Training costs were imputed at $57.8 billion and uncompensated factor services of volunteer workers at $14 billion. (Uncompensated services of military draftees, $16 billion in 1970, were zero in 1976.) We also impute $14 billion for "expense account items of consumption," a figure which with a broader concept than we have been able to apply and better measures might well be, I would judge, considerably higher. Our *TISA GNP*, however, is reduced by subtraction of $171 billion for the value of intermediate product of government already included in output sold by business and government enterprises. We also subtract $35 billion for travel expenses related to work.

Since *TISA* gross national product includes expanded measures of both capital formation and the services of the additional capital, estimates of net national product may offer more meaningful comparisons. Here, we find the *BEA* figure of $1,522 billion just two-thirds of the *TISA* $2,281 billion. The *BEA* national income of $1,360 billion was 66 percent of the corresponding *TISA* net national income (minus net revaluations), which came to $2,062 billion.

Finally, while clearer readings must await completion and refinement of our constant dollar estimates, some patterns may be discerned in relative rates of growth over the thirty-year period from 1946 to 1976 and each of the ten-year subperiods. The *TISA* net national product did seem to grow more rapidly than *BEA* net national product, 8.1 percent as against 7.1 percent. The contrasting comparison of national incomes without net revaluations, however, points up the fact that the difference is largely attributable to major capital gains, particularly in land. The more rapid growth of *TISA* net national income minus net revaluations in the 1966–76 subperiod, 8.7 percent as compared to the *BEA* national income growth rate of 8.1 percent, reflects essentially the greater rates of growth of capital formation related to government and of net intangible capital formation generally.

All this is a report of work in progress. Some major aspects are still conceptually excluded, such as the measure of nonmarket output in the form of environmental protection, except insofar as this is reflected in the value of land and other capital. Our indications of the major magnitudes of investment and output not in conventional measures, however, should offer a strong note of caution to efforts to explain productivity and its growth and decline by relating conventional measures of factor, supply whether of labor or capital, to conventional measures of output.

We may hope as we complete and refine our own constant dollar measures, within the total incomes system of accounts, to offer new light on these old, yet pressing issues.

APPENDIX: *TISA* National Income and Product Account, Preliminary Estimates in Billions of Dollars, 1976

Debits

1. Labor income	1,780.6
1. Compensation of employees	1,035.9
2. Additional imputations	794.4
1. Training costs	57.8
2. Expense account items of consumption	13.6
3. Opportunity costs of self-employed	91.3
4. Opportunity costs of students	203.0
5. Unpaid household work	413.7
3. Less: expenses related to work	34.7
2. Interest	647.9
1. Interest paid	124.7
2. Net imputed interest	517.0
3. Net interest, rest of world	6.2
3. Net operating surplus	−71.4
1. Corporate profits	118.8
2. Private noncorporate income	100.2
3. Business investment in research and development	16.0
4. Government enterprise surpluses	8.1

5. Less: Imputed factor incomes in business	314.5
4. Rental income on owner-occupied nonfarm dwellings	−67.6
1. Gross	11.0
2. Less: Net imputed interest on owner-occupied dwellings and land	78.5
5. Net Revaluations	182.0
1. Land	127.7
2. Owner-occupied dwellings	44.7
3. Other structures and equipment	22.6
4. Consumer durables and semi-durables	−9.6
5. Inventories	−3.4
6. Net surplus (3+5)	110.6
7. National income (1+2+4+6)	2,471.5
8. Less: Intangible capital consumption	227.3
1. Research and development	29.7
2. Human capital	197.7
9. Net national income (7−8)	2,244.2
10. Business transfer payments	30.5
1. Media support	20.8
2. Other	9.8
11. Uncompensated factor services	14.2
12. Net indirect taxes	−15.2
1. Indirect taxes	155.7
2. Less: Intermediate product of government included in sales	107.9
13. Statistical discrepancy	7.0
1. *BEA*	4.2
2. Other	2.8
14. Charges against net national product	2,280.8
15. Capital consumption allowances	656.6
1. Tangible	429.3
2. Intangible	227.3
1. Research and development	29.7
2. Human capital	197.7
16. Charges Against Gross National Product	2,937.4

Credits

17. Consumption	1,628.1
1. Expenditures for services and nondurables	572.0
1. Gross	606.7
2. Less: Expenses related to work	34.7
2. Expense account items of consumption	13.6
3. *BEA* imputations other than housing	35.6
4. Subsidies allocated to consumption	3.4
5. Transfers	158.3
1. From business	15.4
1. Media support	14.6
2. Health and safety	0.7
2. From nonprofit institutions	50.2
3. From government enterprises	2.0
4. From government	90.4
6. Nonmarket services produced in households	845.4
1. Net space rent of owner-occupied nonfarm	96.6
2. Other capital services	342.7
1. Durables	236.0
2. Semidurables	106.0
3. Inventories	0.7
3. Labor services	616.7
4. Less: Services allocated to investment	210.5
18. Gross domestic capital and accumulation	1,301.9
1. Original cost	1,116.1
1. Tangible	558.3
1. Fixed	542.3
1. Business	167.0
1. Nonresidential	156.6
2. Residential (excluding owner-occupied housing)	10.4
2. Nonprofit institutions	5.9
3. Government (including government enterprises)	75.8
1. Structures and equipment	59.6
2. Government product accumulated	16.2
4. Households	293.6
1. Owner-occupied nonfarm dwellings	52.2
2. Consumer durables	156.6

3. Semi-durables	84.8
2. Change in inventories	16.0
1. Business, non-profit and government enterprises	10.2
2. Government	5.1
3. Households	0.6
2. Intangible	557.8
1. Research and development	37.7
2. Education	444.9
3. Health	75.1
2. Subsidies and government enterprise transfers allocated to investment	3.7
3. Net revaluations	182.0
1. Land	127.7
2. Structures and Equipment	67.3
1. Owner-occupied dwellings	44.7
2. Other	22.6
3. Household durables and semidurables	−9.6
4. Inventories	−3.4
19. Net exports	7.4
1. Exports	163.2
2. Imports	155.7
20. Gross National Product	2,937.4

REFERENCES

R. Eisner, "Total Incomes in the United States, 1959 and 1969," *Rev. Income Wealth*, Mar. 1978, *24*, 41–70.

______, "Capital Gains and Income: Real Changes in the Value of Capital in the United States, 1946–77," in Dan Usher, ed., *The Measurement of Capital,* Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 45, Chicago 1980.

______ **and D. H. Nebhut,** "An Extended Measure of Government Product: Preliminary Results for the United States, 1946–76," *Rev. Income Wealth*, 1980 forthcoming.

F. T. Juster, "A Framework for the Measurement of Economic and Social Performance," in Milton Moss, ed., *The Measurement of Economic and Social Performance*, New York 1973.

John W. Kendrick, *The Formation and Stocks of Total Capital*, New York 1976.

Milton Moss, *The Measurement of Economic and Social Performance*, New York 1973.

William Nordhaus and James Tobin, "Is Growth Obsolete?," in their *Economic Growth*, Vol. V, New York 1972.

N. Ruggles and R. Ruggles, "A Proposal for a System of Economic and Social Accounts," in Milton Moss, ed., *The Measurement of Economic and Social Performance*, New York 1973.

Bureau of Economic Analysis, *The National Income and Product Accounts of the United States, 1929–74, Statistical Tables*, Washington 1976.

DISCUSSION

John W. Kendrick, George Washington University and University of Hawaii: I am in full agreement with the central message of Edward Denison's paper that tangible capital formation is only one of a number of important sources of economic growth. In view of the extensive studies of growth and development during the past generation, particularly Denison's own well-known growth accounting exercises, the exclusive focus on investment by some policy makers does seem strange.

Perhaps the most charitable explanation is that these policymakers construe investments in a broader sense, and expect that measures to create a more favorable climate and incentives for tangible investment will also help to stimulate research and development outlays and human investments, particularly in education and training.

The decline in the rate of advance of technological knowledge after 1966, reflecting a sharp drop in the *R&D/GNP* ratios, was undoubtedly a factor reducing the rate of growth in real capital stocks per worker. Although education and training per worker continued to grow at or above prior rates, slowing technological advance appears to have reduced the rate of return on, and productivity of, education and training. Nevertheless, as documented in my monograph *The Formation and Stocks of Total Capital*, the growth of real capital stocks, tangible and intangible, accounted for only about three-fourths of the growth of real *GNP* in the United States, 1929–69. The other quarter was due to noncapital forces such as economies of scale, resource reallocations, the net impact of governmental interventions, and changes in labor efficiency as such.

I believe that William Freund, author of the study *Reaching a Higher Standard of Living* (to which I contributed the analysis of sources of productivity growth, and of the slowdown after 1973) had a broader view of the policies required to return to a growth rate of real product per hour of 3 percent than was implied by Denison. While Freund put primary emphasis on policies to stimulate tangible capital formation, he linked this with encouragement of *R&D* and innovation, and listed six other measures to promote productivity, including improvement of education and training, and relaxation of unnecessary government regulations.

Although Freund did not attempt to quantify the contribution to accelerated productivity growth that might come from various policy options, I attempted such a calculation in my recent paper in *Contemporary Economic Problems, 1979,* edited by William Fellner. There I proposed various policy measures designed to raise the growth of real product per hour in the *U.S.* business economy by 1.3 percentage points, from the 2.1 percent a year average annual rate contained in a *BLS* projection that assumed no new major policy initiatives, up to 3.4 percent a year which is close to the rate achieved in the 1948–66 period.

Tax measures designed to increase the after-tax rate of return on new gross private domestic investment and increase its ratio to *GNP* to 12 percent would contribute 0.3 percentage points to the acceleration, a number consistent with Denison's calculations. Of the remaining 1.0 percentage point acceleration (in total factor productivity), 0.3 was to be achieved by increases in real federal government funding of *R&D* and extending the investment tax credit to cover private *R&D*; plus another 0.1 as a result of faster diffusion of technological knowledge as higher investment reduced the average age of fixed capital goods. Favorable volume changes would contribute 0.2 points as faster growth meant larger economies of scale, and as rates of utilization of capacity were raised. The net impact of government would be 0.2 point more favorable as an effective productivity-enhancing program was introduced, and as the negative impacts of government regulations were reduced. Finally, a stronger growth of education and training per worker would contribute 0.1 percent, and so would an increase in labor efficiency (the ratio of actual to potential under given technology) as a result of encouragement of company productivity im-

provement programs including joint labor-management productivity committees. In another paper devoted solely to policy issues, I have developed more than 100 options for accelerating productivity growth distributed among seven major groups of causal factors in addition to tangible capital formation.

I take issue with Denison's assertion that slower capital formation after 1973 "was mainly a reaction to unemployment and unutilized capacity." I believe that the reduced rates of return on investment in the 1970's compared with the 1960's, after IVA and CCA, and blows to business confidence from inadequate governmental programs to deal with accelerating inflation, excessive regulations, and energy and balance-of-trade problems accounted for the sluggish recovery of investment after 1975, which exacerbated the problems of unemployment, excess capacity, and slow productivity growth.

In conclusion, I wish to emphasize the usefulness of Denison's growth accounting scheme as a framework for macro-economic projections and for developing policy recommendations to accelerate the growth of productivity and real product. His studies of the sources of growth have been generally recognized as of great value, and I would recommend to others the use of his approach not only for projections, but as a means of estimating the effects of various policy prescriptions in the manner which he illustrated in his paper with respect to capital formation. We should not be discouraged, however, when any given major policy measure is estimated to produce growth effects of only 0.1 or 0.2 percentage points. When a number of such complementary policies are pursued, the combined effect can sum to one or more percentage points, which, when compounded, can make a very big difference to the long-run economic outlook of the nation!

OLI HAVRYLYSHYN, George Washington University: With his paper, Robert Eisner has now made available data for 1959, 1969, and 1976 on his Total Income System of Accounts (*TISA*). It is therefore a most appropriate time to commend Eisner for the valuable contribution he has *already* made to the task of extending and modifying the National Accounts beyond the market-output measure of *GNP*. I believe Eisner has completed the first half of his contribution: the massive effort of data gathering, collating, and estimating. I look forward to the second half of his contribution in the future: applications of *TISA* to various economic problems, of which this paper is but a precursor.

The focus of this session being capital formation and growth problems, let me concentrate on the directions one may take in applying *TISA* to this issue. Before that, I have a general remark on *TISA*'s methodology valuation. It seems necessary now to confront Eisner's data with other estimates in the Measure of Economic Welfare (*MEW*) spirit. Thus, for example, why is the *TISA* estimate of "household-work," at about 28–30 percent of *GNP*, somewhat lower than that of most others at 33–35 percent? Why is his measure of "media contributions" at $20.8 billion higher than that of Jack Cremeans' at $16.8 billion? The newly started Measures of Economic Well-Being Branch at *BEA* is providing us with a number of such estimates to add to other earlier ones (Kendrick, Ruggles and Ruggles, Havrylyshyn), all of which together are a basis for comparison with *TISA*.

With regard to applications of *TISA* to the growth problem, three principal elements stand out: 1) the volume and movements of total capital formation differs from the usual *BEA* data; 2) *TISA* provides a new measure of capital: education; and 3) *TISA* revalues capital stock annually.

On the first, *TISA*'s capital formation total is far greater than *BEA*'s, the latter being a mere 22 percent of *TISA*'s; even excluding intangible investment the figure is only 30 percent. One may therefore ask whether, by underestimating investment in the *BEA* measure, one does not underestimate the role of capital in growth through Denison-like analysis? On the point of productivity decline in the 1970's, Eisner's *TISA* also tells us that the growth rates of government investment have declined con-

sistently from 9.1 percent in 1951–55 to 3.8 percent in 1971–76, a pattern coinciding with productivity. It is easy to tell a plausible story linking the two: a direction *TISA* applications should take.

However, one must be careful in so doing to extract from *TISA* components that are pertinent to productivity. Specifically, one should exclude residential and durables investments, which in the case of tangibles gives a 1976 value of $248.2 billion, not much different from the $243.0 billion normally defined as investment. Only upon adding intangibles, in particular education, do we obtain a very different value of $821.7 billion.

On the second point, human capital, *TISA*'s numbers are of great relevance. It is not, however, that we haven't known about human capital, or not applied it in analysis, or not had earlier estimates (for example, Kendrick, Graham and Webb). Rather, what matters is that *TISA* systematically provides these values in exact parallel to physical capital values, and now *both* can be used directly in sources of growth analysis, or aggregate production function models. Note that the magnitude is substantial. In Table 1, human capital formation inclusive of health is $535 billion, equal to total tangible investment, and *twice* the value of *BEA* investment. Note further that its growth rate has been increasing: we should ask if this is a positive sign for a return to higher productivity rates.

The third point concerns revaluations with which I have some problems. Revaluation results in a higher growth rate for *TISA* than for *GNP* in the period 1946–76, of 8.1 vs. 7.1 percent. Let me argue these are not real current output gains, but at best reflections of past and perceived future gains in wealth. In the latter sense this is an important variable to include in a permanent income analysis of consumer behavior, and may cast light on observed declines in savings rates, and even on inflationary trends. However, being a measure of income and welfare does not necessarily mean it is a measure of output of capital: revaluing higher-priced land or a metal press is indeed a personal gain, but it differs from a new metal press in the latter's potential to produce more output. This distinction is especially important to issues of productivity. Thus, care must be taken to choose a meaning of capital in *TISA*; one should go beyond *BEA* to include tangible government investment, and human capital, but one should stop short of *TISA*'s total definition by excluding revaluations, household durables, and residential investment.

To conclude, let me concur with Eisner as to *TISA*'s potential to "offer new light on these old, yet pressing issues." However, let me add a note of caution. *TISA* is not a replacement for *GNP*, but a needed and useful complement to be used with great care.

INTEWAR MACROECONOMICS FROM THE PERSPECTIVE OF THE 1970'S

Exchange Rates, Prices, and Money: Lessons from the 1920's

By JACOB A. FRENKEL*

The experience with flexible exchange rates during the 1920's has proven to be extremely important in shaping our current thinking about a variety of issues including the choice among alternative exchange rate regimes, the role of speculation in the market for foreign exchange, the purchasing power parity doctrine, and the determinants of equilibrium exchange rates.

Probably no event in monetary history has been studied more closely than the German hyperinflation. Economists have been attracted to study this episode since it provides an environment that is close to a controlled experiment which is so rare in the study of social sciences. It also provides a convenient starting point for the reexamination of theories in circumstances in which the predominant disturbance is of a monetary origin. However, interest in the experience with flexible exchange rates during the 1920's is not confined only to the lessons from the German hyperinflation. From the viewpoint of economic research, that experience provides also the opportunity to conduct a comparative study of the operation of flexible exchange rates under "normal" conditions. Specifically, until the return to gold by Britain (in 1925), many countries adopted a flexible exchange rate system. This system was successful in insulating most of the world from the direct consequences of the extraordinary German hyperinflation of 1921–23. Thus, during the same period in which Germany was experiencing the hyperinflation, much of the rest of the world was operating under practically "normal" conditions.

This paper summarizes the results of an empirical study of the operation of flexible exchange rates during the 1920's under both the hyperinflationary conditions (based on the experience of Germany) and under the normal conditions (based on the experience of Britain, the United States and France). Section I deals with some general characteristics of the market for foreign exchange by examining the relationship between spot and forward exchange rates. Section II deals with the relationship between exchange rates and prices by examining aspects of the purchasing power parity doctrine. Section III deals with the determinants of exchange rates within the context of a simple monetary model.

I. Characteristics of the Market for Foreign Exchange

This section examines the relationship between spot and forward exchange rates. The hyperinflation period is analyzed using the Mark/Pound exchange rate during February 1921–August 1923, and the normal period is analyzed using the Franc/Pound, the Dollar/Pound, and the Franc/Dollar exchange rates (of which only two are independent) during the period February 1921–May 1925. The length of the hyperinflation period was determined by the availability of data on the forward exchange rate, whereas the normal period was

*Professor of economics, University of Chicago and National Bureau of Economic Research, Inc. I am indebted to Lauren Feinstone and Craig Hakkio for efficient research assistance and to the National Science Foundation grant SOC 78-14480 for financial support. This research is part of the NBER's Program in International Studies. The views expressed are my own and not necessarily those of the NBER.

terminated by the return of Britain to gold. Data sources for the hyperinflation are the same as in my 1977 article, and the sources for the normal period are the same as in my 1978 article, and my forthcoming article with Kenneth Clements. All data are monthly.

To examine the relationship between spot and forward rates, I first regress the logarithm of the current spot exchange rate, $\ln S_t$, on the logarithm of the one-month forward exchange rate prevailing at the previous month, $\ln F_{t-1}$, as in equation (1).

$$(1) \qquad \ln S_t = a + b \ln F_{t-1} + u_t$$

If the forward exchange rate is an unbiased forecast of the future spot exchange rate, the constant terms in equation (1) should not differ significantly from zero,[1] and the slope coefficient should not differ significantly from unity. Efficiency of the market requires that the residuals from the estimated regression be serially uncorrelated. Table 1 contains the ordinary least squares (OLS) estimates of equation (1). As may be seen for the normal period, the hypotheses that the constant terms do not differ significantly from zero and that the slope coefficients do not differ significantly from unity cannot be rejected (at the 95 percent confidence level). The Durbin-Watson ($D.W.$) statistics indicate the absence of first-order autocorrelated residuals; an examination of higher order correlations (up to 12 lags) shows that no correlation of any order is significant. For the hyperinflation case, the constant term is somewhat negative while the slope coefficient is somewhat above unity (at the 95 percent confidence level). Most importantly, however, the residuals from the equations are serially uncorrelated (up to 12 lags). In all cases the inclusion of an additional lagged value of the forward rate does not lower the standard error of the regressions, and its coefficient does not differ significantly from zero. This finding is consistent with the hypothesis that the forward rate prevailing at period $t-1$ summarizes all available information and thus the inclusion of data from period $t-2$ does not improve the fit.

The efficiency of the market can also be analyzed from a different angle. Consider equation (2):

$$(2) \qquad x_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{n} \beta_i x_{t-i} + \gamma \pi_{t-1} + w_t$$

where x_t denotes the percentage change in the spot exchange rate ($\ln S_t - \ln S_{t-1}$), π_{t-1} denotes the forward premium on foreign exchange ($\ln F_{t-1} - \ln S_{t-1}$), t denotes time, n denotes the number of lags, and w is an error term. If the forward premium on forward exchange summarizes all available information concerning the future evolution of the exchange rate, then given the value of the forward premium, the past history of the rate of depreciation should not "help" the prediction, and the joint hypotheses that α_1 and β_i are zero should not be rejected. Applying these tests (using 4, 5, and 6 lags) reveal that for the German hyperinflation the null hypothesis can not be rejected at the 95 percent confidence level. For the normal period, the results are similar but somewhat less strong: in some cases the null hypothesis is rejected (marginally) at the 95 percent confidence level, but not at the 99 percent level. Similar results are obtained from testing the joint hypotheses that α_1 and β_i are zero and that γ is unity. It is concluded therefore that the forward premium on foreign exchange incorporates the relevant available information contained in past exchange rate changes. It should be noted that if we regard the forward premium as a measure of the predicted change in the exchange rate, then this prediction accounts for a very small fraction of the actual variations and thus indicates that the bulk of exchange rate changes are due to "new information." This fact seems to be an empirical regularity which is found in both the hyperinflation period as well as in the normal period.

The assumption underlying equation (1) was that the forward rate $\ln F_{t-1}$ measures

[1]More precisely, if the forward rate is an unbiased forecast of the future spot rate, then the constant term in equation (1) should equal $-0.5\sigma_u^2$; see my 1979 article.

TABLE 1—EFFICIENCY OF FOREIGN EXCHANGE MARKETS: MONTHLY DATA
(Standard Errors in Parentheses)

Dependent Variable: $ln\, S_t$	Constant	$ln\, F_{t-1}$	$ln\, F_{t-2}$	$S.E.$	R^2	$D.W.$	m
Mark/Pound	−.454	1.094		.46	.98	1.89	.38
(Feb. 1921–Aug. 1923)	(.254)	(.029)					
	−.437	1.120	−.029	.47	.98	1.95	
	(.302)	(.187)	(.209)				
Franc/Pound	.169	.962		.07	.91	1.92	.09
(Feb. 1921–May 1925)	(.179)	(.042)					
	.177	.992	−.032	.07	.91	1.97	
	(.187)	(.144)	(.147)				
Dollar/Pound	.057	.964		.02	.93	1.54	.02
(Feb. 1921–May 1925)	(.056)	(.038)					
	.073	1.181	−.229	.02	.93	2.11	
	(.057)	(.143)	(.142)				
Franc/Dollar	.203	.928		.08	.85	1.95	.52
(Feb. 1921–May 1925)	(.149)	(.054)					
	.206	.945	−.018	.08	.85	1.98	
	(.156)	(.145)	(.146)				

Note: $S.E.$ is the standard error of the equation; the m-statistic tests for the absence of errors in variables arising from using $ln\, F_{t-1}$ as a proxy for the expected future spot rate. It is distributed χ^2 with 2 degrees of freedom. To compute the m-statistic, the equations were reestimated using lagged values of the dependent and independent variables, a constant, time, and time squared as instruments.

the *expected* spot rate for period t. If, however, the forward rate is a "noisy" proxy for the expected future value of the spot rate (i.e., it measures it with a random error), then the assumption that $ln\, F_{t-1}$ and u_t in equation (1) are independent would be violated and the *OLS* procedure would yield biased estimates due to the classical errors in variables bias. To examine this possibility I follow the specification test outlined by Jerry Hausman. The m-statistics reported in Table 1 indicate that in all cases the use of the forward exchange rate as a proxy for expectations does not introduce significant errors in variables bias.

The principal conclusions that may be drawn are that the behavior of the foreign exchange market during the normal period was consistent with the general implications of the efficient market hypothesis, and that the forward exchange rate summarized the relevant available information concerning the future evolution of the rate. These results are somewhat less clear for the hyperinflation period where it is seen (in Table 1) that the forward rate seems to have underpredicted the future spot rate. However, in view of the unprecedented acceleration of the rates of inflation and depreciation, one could expect that, while individuals learn the new structure, mistakes would be made and expectations would initially underpredict the actual course of events. With this perspective I interpret the overall behavior of the foreign exchange market during the hyperinflation as being in accord with the broad principles of an efficiently functioning market.

II. Exchange Rates and Prices

The relationship between exchange rates and prices is summarized by the purchasing power parity (*PPP*) doctrine which states that the equilibrium exchange rate between domestic and foreign currencies equals the ratio between domestic and foreign prices.

In its general form the *PPP* relationship can be written as

$$(3) \quad ln\, S_t = a + b\, ln\, P_t - b^*\, ln\, P_t^* + u_t$$

where S_t, P_t and P_t^* denote, respectively, the exchange rate, domestic and foreign price indices (with an asterisk denoting quantities

TABLE 2—PURCHASING POWER PARITIES: MONTHLY DATA
(Standard Errors in Parentheses)

Dependent Variable: $ln\,S_t$	Price Index	Constant	$ln(P/P^*)$	Test Restriction $F(2,48)$	$S.E.$	R^2	$D.W.$	ρ
Mark/Pound (Feb. 1921–Aug. 1923)	wholesale	−1.676 (.178)	1.026 (.017)	–	.221	.99	2.01	.24
	cost of living	−1.575 (.423)	1.084 (.041)	–	.367	.99	2.06	.50
Franc/Pound (Feb. 1921–May 1925)	wholesale	.562 (.207)	1.141 (.064)	4.97	.044	.97	1.82	.53
	material	.613 (.180)	1.081 (.054)	4.67	.042	.97	2.18	.48
Dollar/Pound (Feb. 1921–May 1925)	wholesale	−.118 (.482)	.897 (.267)	.53	.019	.94	1.99	.85
	material	−.073 (.453)	.847 (.245)	11.11	.022	.91	1.83	.80
Franc/Dollar (Feb. 1921–May 1925)	wholesale	1.183 (.157)	1.091 (.109)	2.32	.054	.93	1.70	.58
	material	1.243 (.130)	.992 (.085)	2.14	.050	.94	1.74	.54

Note: $S.E.$ is the standard error of the regression. R^2 was computed as $1-var(\hat{u}_t)/var(ln\,S_t)$. ρ is the final value of the autocorrelation coefficient; an iterative Cochrane-Orcutt transformation was employed. Two-stage least squares estimation was used following Ray Fair's method with lagged values of the dependent and independent variables, a constant, time and time squared as instruments. The F-statistics test the validity of the joint restrictions that domestic and foreign prices have the same coefficients and that the elasticity is unity. Critical values for $F(2,48)$ are 3.19 (95 percent) and 5.08 (99 percent).

pertaining to the foreign country), and where u_t denotes an error term.

From the empirical viewpoint, several issues may be raised: (i) What price index should be used in equation (3)? (ii) Are the coefficients of domestic and foreign prices the same? (iii) Are the data consistent with the hypothesis that the coefficients of domestic and foreign prices are both unity; i.e., are $b=b^*=1$?

These questions are examined below for both periods using alternative price indices. To allow for a simultaneous determination of prices and exchange rates, equation (3) was estimated using a two-stage least squares estimation procedure. The estimates are reported in Table 2, as are the values of F-statistics relevant for testing the hypothesis that $b=b^*=1$.

On the whole, for the normal period the results show that, except for one case (corresponding to the regression of the Dollar/Pound exchange rate on the ratio of the material prices indices), the data are consistent with the joint restrictions implied by the homogeneity postulate. For example, the elasticity of the Franc/Dollar exchange rate with respect to the wholesale price ratio is 1.091 and the joint hypothesis that $b=b^*=1$ is not rejected since the value of the F-statistic is 2.32 while the critical value (at the 95 percent confidence level) is 3.19.

In applying equation (3) to the hyperinflation period, it was assumed that the variations in P/P^* were completely dominated by variations in German prices so that the foreign price could be viewed as being fixed. As is seen in Table 2, the PPP doctrine seems to be satisfied for the wholesale price index where the elasticity of the exchange rate with respect to that price is 1.026. It holds less well for the cost-of-living index which includes prices of many nontraded goods. Nonetheless, even for the cost-of-living index the elasticity is not too far from unity.

In assessing these results it is important to recall that the estimates are based on monthly data, and that short-run deviations from *PPP* may reflect the fact that not all markets adjust at the same speed. It is also important to recognize that the *PPP* doctrine describes an equilibrium relationship between two endogenous variables. As such, it should be viewed as a shortcut rather than a substitute for a more complete model of the determination of prices and exchange rates. The main usefulness of the doctrine is in providing a guide as to the general trend of exchange rates rather than the day-to-day fluctuations. The main lesson for policy that is provided by the evidence in Table 2 is that the exchange rate and the price level cannot be divorced from each other and, as a first approximation, policies which affect the trend of domestic (relative to foreign) prices also affect the exchange rate in the same manner.

III. Exchange Rates, Money, and Expectations

This section analyzes the determinants of exchange rates during the 1920's from the perspective of the monetary (or the asset market) approach to the exchange rate. Being a relative price of two assets (moneys), the equilibrium exchange rate is attained when the existing *stocks* of the two moneys are willingly held. Therefore, it is convenient (though not necessary) to analyze the determinants of the exchange rate in terms of the supply and the demand for these moneys.[2] The central insight obtained from this approach is the recognition that expectations concerning the future exchange rates are among the prime determinants of current exchange rates.

A simple exchange rate equation which incorporates these considerations is specified in equation (4):

$$(4) \qquad S=f(M,M^*,y,y^*,\pi)$$

where M and M^* denote domestic and foreign money supplies, y and y^* denote domestic and foreign incomes, π denotes the forward premium on foreign exchange (which is assumed to measure the expected depreciation of the currency), and S denotes the exchange rate (i.e., the price of foreign exchange in terms of domestic currency). While this equation can be derived from a specific monetary model (which allows for real and monetary factors), a similar set of variables would also appear in the reduced form of a variety of alternative models (see Rudiger Dornbusch). The predictions are that, *ceteris paribus*, (i) a rise in the supply of domestic money will raise S (i.e., depreciate the home currency) while a rise in the supply of foreign money will lower S. The homogeneity postulate requires that the elasticity of S with respect to M and M^* be unity and minus unity, respectively; (ii) a rise in domestic income due, for example, to a rise in productivity which raises the relative demand for domestic money, will appreciate the currency (lower S), while a rise in foreign income (which raises the relative demand for foreign money) will depreciate the currency,[3] and (iii) a rise in the forward premium on foreign exchange will lower the relative demand for domestic money and will depreciate the currency (raise S).

Estimates of equation (4) are reported in Table 3. For the hyperinflation period it was assumed that variations in the ratio of the two money supplies were dominated by variations in the German money supply and that variations in the demands for moneys were dominated by changes in expectations so that changes in incomes and in the for-

[2]This emphasis on the stock demand and supply of moneys was also adopted by John Maynard Keynes in explaining the value of the French Franc: "What, then, has determined and will determine the value of the Franc? First, the quantity, present and prospective, of francs in circulation. Second, the amount of purchasing power which it suits the public to hold in this shape" (introduction to French edition 1924, p. xviii). The renewed emphasis on stock considerations should not be interpreted to imply that flow relationships are not important; the equilibrium is attained by the interaction of flow and stock equilibrium conditions.

[3]When one of the currencies is a reserve currency or when individuals and firms hold portfolios of many currencies the effects of changes in incomes on the *relative* demands for moneys (and thereby on the exchange rate) are not clear-cut.

TABLE 3—EXCHANGE RATE EQUATIONS: MONTHLY DATA
(Standard Errors in Parentheses)

Dependent Variable: $\ln S_t$	Constant	$\ln M$	$\ln M^*$	$\ln y$	$\ln y^*$	π	$S.E.$	R^2	$D.W.$	ρ
Mark/Pound	−6.030	.970				3.886	.340	.99	2.56	.89
(Feb. 1921–Aug. 1923)	(1.696)	(.092)				(1.131)				
Franc/Pound	.001	.999	−.972	.188	.926	3.914	.069	.92	1.86	1.00
(Feb. 1921–May 1925)	(.010)	(.099)	(.099)	(.281)	(.520)	(.970)				
Franc/Dollar	.006	.995	−.995	.225	−.369	3.971	.075	.86	1.81	1.00
(Feb. 1921–May 1925)	(.011)	(.099)	(.100)	(.327)	(.370)	(.974)				

Note: The Mark/Pound exchange rate equation was estimated using two-stage least squares following Fair's method with lagged values of the dependent and independent variables, a constant, time, and time squared as instruments. ρ is the final value of the autocorrelation coefficient; an iterative Cochrane-Orcutt transformation was employed. R^2 was computed as $1 - var(\hat{u}_t)/var(\ln S_t)$. The Franc/Pound and the Franc/Dollar exchange rate equations were estimated in first difference form using the Theil-Goldberger mixed-estimation procedure with stochastic restrictions.

eign money supply could be ignored. As is evident, the results are fully consistent with the prior expectations. The elasticity of the exchange rate with respect to the money stock does not differ significantly from unity while the (semi) elasticity with respect to the forward premium is positive indicating that an expected future depreciation of the currency results in an immediate depreciation. The order of magnitude of the latter (semi) elasticity is similar (in absolute value) to the interest (semi) elasticity of the demand for money.

The results for the normal period are less definitive. The *OLS* estimates of equation (4) (not reported here) are extremely imprecise. Due to the limited degree of variation in the series and the high degree of collinearity, the information that is contained in the sample is not sufficient to provide precise estimates of the various parameters. To obtain more precise estimates, the sample information was supplemented by prior information about the elasticities with respect to the money supplies, and the forward premium (which was based on the estimates for the hyperinflation period) according to the Theil-Goldberger mixed-estimation procedure. This procedure was adopted for the Franc/Pound and the Franc/Dollar exchange rates after verifying (according to a χ^2 test) that the sample and the prior information were compatible with each other; the Dollar/Pound exchange rate failed the χ^2 test.[4] As is evident from Table 3, the results are consistent with the predictions. For both exchange rates, the elasticity with respect to domestic and foreign money supplies are unity and minus unity, respectively, and the (semi) elasticities with respect to the forward premium are positive. The income elasticities are not significant (at the 95 percent confidence level). In the only case where the income elasticity comes close to being significant (foreign income in the Franc/Pound exchange rate), it has the correct positive sign. While the results for the normal period seem to be consistent with the theoretical predictions, the fact that the sample provides very little information suggests that the data would also be compatible with different values of the prior information and that the test is not sufficiently powerful to discriminate among alternative approaches.

[4]For details see my forthcoming article with Clements. For simplicity equation (4) does not include the relative price of traded to nontraded goods as an explanatory variable; this might be responsible for the results for the Dollar/Pound rate. Clements and I show that this relative price played an important role in determining the Dollar/Pound rate. Another useful modification of equation (4) would recognize that the exchange rate effects of anticipated changes in money and income are fundamentally different from the effects of unanticipated changes.

IV. Concluding Remarks

The experience of the 1920's provides the opportunity to conduct a comparative study of the behavior of a flexible exchange rates system under "normal" conditions as well as under highly inflationary conditions. The three issues examined in this paper are the relationship between spot and forward exchange rates, the relationship between exchange rates and prices, and the relationship between exchange rates, money, and expectations.

In a recent paper dealing with inflation and unemployment Robert Gordon analyzed the reactions of a modern-day Rip Van Winkle who had become well acquainted with the earlier literature but who only recently awoke from a decade-long nap. It is interesting to examine a related experiment: suppose that Rip who was well acquainted with the data reported in this paper went to sleep in 1925 to be awoken in the 1970's. Would his human capital of the 1920's vintage be obsolete? This question is of special interest since world capital markets have become much more integrated, the role of "real" shocks in the 1970's have become much more important, views about the role of government in the conduct of macro-economic policies have changed, the roles of tariff and non-tariff barriers to trade as well as the degree of exchange rate management have been altered, and finally, the International Monetary Fund has been created.

A comparison of the lessons from the 1920's with the experience in the 1970's suggests that the answer is somewhat mixed. The various characteristics of the relationship between spot and forward exchange rates which were discussed in Section I seem to be the most robust. It seems to be stable between the normal and the hyperinflation periods of the 1920's, and the evidence from the 1970's suggests that it is also stable over time. The major difference relates to the relationship between exchange rates and prices: while the *PPP* doctrine held up reasonably well during the 1920's, the evidence from the 1970's indicates large departures from *PPP*. These departures have not yet been fully explained. Finally, the relationship between exchange rates, money, and the forward premium was shown to be very significant during the hyperinflation but less so during the "normal" period. A similar pattern seems to reemerge in the 1970's where analogous exchange rate equations are more successful in circumstances with high inflation than in circumstances with low inflation.

REFERENCES

J. F. O. Bilson, "Rational Expectations and the Exchange Rate," in Jacob A. Frenkel and Harry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*, Reading 1978.

K. W. Clements and J. A. Frenkel, "Exchange Rates, Money and Relative Prices: The Dollar-Pound in the 1920's," *J. Int. Econ.*, 1980, forthcoming.

R. Dornbusch, "Monetary Policy Under Exchange Rate Flexibility," in *Managed Exchange Rate Flexibility: The Recent Experience*, Fed. Reserve Bank Boston, *Monetary Conference Series*, no. 20, 1978.

Jacob A. Frenkel, "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence," *Scand. J. Econ.* May 1976, *78*, 200–24; reprinted in his and Harry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*, Reading 1978.

______, "The Forward Exchange Rate, Expectations and the Demand for Money: The German Hyperinflation," *Amer. Econ. Rev.*, Sept. 1977, *67*, 653–70.

______, "Purchasing Power Parity: Doctrinal Perspective and Evidence from the 1920's," *J. Int. Econ.*, May 1978, *8*, 169–91.

______, "Further Evidence on Expectations and the Demand for Money During the German Hyperinflation," *J. Monet. Econ.*, Feb. 1979, *5*, 81–96.

______, (1980a) "Flexible Exchange Rates in the 1970's" in *Stabilization Policies: Lessons from the 1970's and Implications for the 1980's*, Center for the Study of

American Business, Washington Univ., St. Louis 1980.

______, (1980b) "The Forward Premium on Foreign Exchange and Currency Depreciation During the German Hyperinflation: Reply," *Amer. Econ. Rev.*, 1980 forthcoming.

______ **and K. W. Clements,** "Exchange Rates in the 1920's: A Monetary Approach," in M. June Flanders and Assaf Razin, eds., *Development in an Inflationary World*, New York 1980.

______ **and Harry G. Johnson,** *The Economics of Exchange Rates: Selected Studies*, Reading 1978.

______ **and M. L. Mussa,** "The Efficiency of Foreign Exchange Markets and Measures of Turbulence," *Amer. Econ. Rev. Proc.*, May 1980, *70*, 374–81.

R. J. Gordon, "Recent Developments in the Theory of Inflation and Unemployment," *J. Monet. Econ.*, Aug. 1976, *2*, 185–219.

J. A. Hausman, "Specification Tests in Econometrics," *Econometrica*, Nov. 1978, *46*, 1251–71.

John Maynard Keynes, *A Tract on Monetary Reform*, 1st ed. 1923, French Edition 1924; Vol. IV in *The Collected Writings of J. M. Keynes*, London 1971.

P. R. Krugman, "Purchasing Power Parity and Exchange Rates," *J. Int. Econ.*, Nov. 1978, *8*, 397–407.

M. L. Mussa, "Empirical Regularities in the Behavior of Exchange Rates and Theories of the Foreign Exchange Market," in Karl Brunner and Alan Meltzer, eds., *Policies for Employment, Prices, and Exchange Rates*, Vol. 11, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1979, 9–57.

A Consistent Characterization of a Near-Century of Price Behavior

By ROBERT J. GORDON*

This paper demonstrates that the commonly used Expectational Phillips Curve (*EPC*) framework cannot explain the last eighty-seven years of aggregate price behavior in the United States. The *EPC* explanation, which in its most general form relates price change to expected inflation and the *level* of detrended output, obscures the fact that price change has been much more closely related to the contemporaneous *rate of change* of detrended output. Over the near-century of annual data studied here, a change in output has shown a remarkably consistent tendency to be associated in annual data with a simultaneous change in the price level of about one-half as much. Stated another way, nominal *GNP* changes have been divided consistently, with two-thirds taking the form of output change and the remaining one-third the form of price change. This finding applies not only over the entire 1890–1978 sample period, but also over three subperiods (1890–1929, 1929–53, and 1953–78).

The dominance of the output rate-of-change (*ROC*) explanation of price change over the "level" *EPC* explanation confirms previous findings by Allan Meltzer, myself, and others.[1] Nevertheless much recent literature on both postwar and historical price behavior has shown no awareness of the importance of the *ROC* phenomenon and has continued to specify equations based on the unadorned *EPC*.[2]

A second theme of this paper is that almost a century of price behavior can be explained with a single equation, in contrast to other historical studies that find it necessary to exclude years of depression, war, or both. While the single equation estimated here has a coefficient on the "*ROC* effect" that is extremely stable over the entire sample period, it verifies a marked shift after the Korean War in the formation of expectations regarding the price level and its rate of change. This shift reinforces the emphasis placed by Meltzer and Benjamin Klein on the contrast between the regressive expectations appropriate under a gold standard and the extrapolative expectations used to predict inflation under the postwar fiat money standard. And although this shift is consistent in overall timing with the research of Phillip Cagan and Jeffrey Sachs, its representation differs here because, while the formation of inflation expectations has shifted in the postwar years, the cyclical impact of detrended output changes has not.

*Professor of economics, Northwestern University, and research associate, National Bureau of Economic Research. I am grateful to Allan Meltzer and Jeffrey Sachs for helpful discussions, and to Jon Frye and Ross Newman for their skilled assistance.

[1]Meltzer provides a convincing demonstration of the importance of *ROC* phenomena, and the unimportance of Phillips curve variables, for a sample period that includes sixty-one of the eighty-seven years studied here. R. A. Gordon shows the importance of the change of the unemployment rate in wage equations and reviews the 1958–72 literature, including several earlier articles that included the change in unemployment. My own previous research on postwar structural price equations has always found strong effects of the rate of change of detrended output or a similar demand variable. For examples of such equations, and for evidence that the *ROC* explanation is also important for postwar wage behavior, see my 1977 article, especially pp. 269–70. I discovered the importance of the *ROC* explanation of interwar price behavior in writing a textbook case study (1978, pp. 162–65), and my article with James Wilcox provided subsequent econometric support.

[2]The recent study of historical price behavior by Jeffrey Sachs neither cites any of the papers listed in fn. 1 nor makes any mention of *ROC* variables. Edward Gramlich utilizes a so-called "mainline" model for the explanation of postwar inflation in which neither the wage nor price equations contain any *ROC* variables. Many other similar econometric models of the postwar inflation process could be cited.

I. Specifying a Reduced-Form Price Equation

This paper concentrates on results for annual price data.[3] Several recent papers take as their point of departure a standard *EPC* framework that makes the inflation rate (p_t) depend on the expected inflation rate (p_t^e) and the "gap" between actual and "natural" output (or unemployment). In the following discussion it is convenient to designate *logs* of levels of variables by uppercase letters; rates of change by lowercase letters; and the gap variable as the *log* of the ratio of actual real *GNP* to natural real *GNP* ($\hat{Q}=Q-Q^*$). Thus, the *EPC* hypothesis becomes

$$(1) \qquad p_t=\alpha_0+\alpha_1\hat{Q}_t+\alpha_2 p_t^e+\eta_t \qquad \alpha_1>0$$

where η_t is an error term. We would expect $\alpha_2=1$ if the "natural rate hypothesis" holds and if expected inflation is measured accurately, and $\alpha_0=0$ if the *log* of natural output (Q^*) is measured accurately.

We shall find, however, that (1) is too restrictive a hypothesis to allow adequate characterization of secular changes in *U.S.* price behavior. Instead, a more general approach can be developed if we start with a simple aggregate supply function that allows the difference between the actual and expected price *level* to respond positively to the output ratio ($\hat{Q}$):

$$(2) \qquad P_t=\beta\hat{Q}_t+P_t^e+\epsilon_t \qquad \beta>0$$

The positive slope of the aggregate supply curve can be explained in traditional textbook fashion as resulting from the diminishing marginal productivity and increasing supply schedules of factor inputs, including both materials and labor. It is also consistent with the strong effect of output-ratio variables on the price of inelastically supplied primary products and on oligopolistic price markups found in recent studies of postwar data.[4]

A general specification of the formation of expectations allows agents to distinguish between expected inflation (p_t^e) and shifts in the expected price level (P_t^e):

$$(3) \qquad P_t^e=p_t^e+\lambda P_{t-1}+(1-\lambda)P_{t-1}^e \qquad 0<\lambda<1$$

The expected rate of inflation in turn can be specified to depend on past values of actual and expected inflation and any other relevant information, for example, the past rates of change of nominal or real *GNP*, the money supply, wages, or unemployment.

An appealing feature of (3) is its applicability to price series with differing patterns of serial correlation. For instance, under a gold standard we might find the price level jumping up and down around a stable or slowly moving trend, and in this case the parameter λ might be small. But under the postwar fiat money standard, few increases in the price level have been reversed. Thus rational agents might set $\lambda=1$ and then apply this period's expected inflation rate (p_t^e) to last period's *actual* price level.

> If, for example, an eight percent inflation rate occurred in a particular year [under the gold standard], an inflation rate of approximately minus eight percent would likely occur a short time later.... The gold standard can be considered to have been a period of *mean reversion* in the rate of price change while the current period is one of persistence or long-term *mean revision* in the rate of price change [Klein, pp. 193–94, his emphasis].

When (3) is substituted into (2), with a bit of manipulation we can derive an equation that relates the rate of inflation to the expected rate of inflation, and to both the

[3]Results for annual wage data are briefly discussed below. Quarterly data extending back to 1900 have also been prepared for most of the variables and will be studied in a future paper.

[4]In addition to my 1977 evidence supporting a relation between the price markup and detrended output, Richard Cooper and Robert Lawrence find a significant relation between the prices of nonferrous metals, primary fibers, and raw agricultural products on the one hand, and detrended world industrial output on the other.

level and rate of change of the output ratio:

$$P_t=\beta(\hat{q}_t+\lambda\hat{Q}_{t-1})+p_t^e+\mu_t \tag{4}$$

Note that (4) reduces to the traditional *EPC* equation (1) only if $\lambda=1$.[5]

If the rate of change of nominal *GNP* relative to natural output growth ($\hat{y}_t=y_t-q_t^*$) is exogenous, then the $\hat{q}_t$ term in (4) will be negatively correlated with the error term, given the identity:

$$\hat{q}_t\equiv\hat{y}_t-p_t \tag{5}$$

This problem can be avoided if we substitute (5) into (4) and obtain the following expression that relates the inflation rate to "adjusted" nominal *GNP* growth, the level of the output ratio, and the rate of expected inflation:

$$p_t=\frac{1}{1+\beta}\left[\beta(\hat{y}_t+\lambda\hat{Q}_{t-1})+p_t^e+\mu_t\right] \tag{6}$$

II. Data and Specification

Published annual data series exist for the 1890–1978 period for nominal and real *GNP*, the *GNP* deflator, consumer prices, average hourly earnings, the money supply, and the unemployment rate.[6] The "natural unemployment rate" is assumed to be constant before 1955 for an unemployment concept that excludes self-employed farmers and proprietors, since the latter experience little unemployment, and "natural real *GNP*" (Q^*) is an estimate of the real *GNP* that the economy can produce when operating at its natural rate of unemployment.[7]

Four issues must be discussed before equation (6) can be estimated. The first and most important is the choice of a proxy for the expected rate of inflation(p_t^e). Edgar Feige and Douglas Pearce have emphasized that inflation expectations should be based on all available information making a marginal contribution to the prediction of inflation that is worth more than its marginal acquisition cost. Because of the shift in monetary standards over our sample period, it is important that we allow agents to shift the variables used to form expectations, and the coefficients applied to those variables.[8] We shall find, for instance, that the lagged inflation rate was very important in helping to predict the persistence of inflation observed in the post-1952 period but was of no use in predicting inflation during the gold standard era. We also allow lagged values of the rate of change of nominal income and the money supply to influence the formation of expectations. The presence of the lagged output ratio ($\hat{Q}_{t-1}$) in equation (6) introduces an identification problem, since we do not know whether a positive coefficient indicates an important influence of that ratio on expected inflation, a high value of λ, or some mixture of the two.[9] A final difficulty is that agents observe economic data at shorter frequencies than a year, so that current variables may contain information used to form expectations. It is assumed here that expectations are based entirely on last year's data, but in future research quarterly versions of these equations will be estimated.

[5]Sachs rules out any structural relation between the inflation rate and the change in the output ratio by assumption. If the price markup in his equation (8) is allowed to be a function of the level of the output ratio, then his model becomes underidentified. See also fn 15.

[6]The starting date of 1890 is chosen because it marks the beginning of the annual Kendrick *GNP* series, Rees' cost-of-living and average hourly earnings series, and the Lebergott unemployment series. A data appendix is available for interested readers. Equations have been run with consumer prices as an alternative dependent variable and with the unemployment gap replacing the output ratio, and there was no important difference from the results presented in Table 1.

[7]The methodology is described in my book, Appendix C.

[8]Thus this paper recognizes Thomas Sargent's criticism that earlier Phillips curve work did not take account of the actual process by which inflation is generated.

[9]If a short-frequency business cycle were expected, the effect of $\hat{Q}_{t-1}$ on expectations might be negative, raising the possibility that our estimates of the coefficient on $\hat{Q}_{t-1}$ might be biased toward zero. This might help to explain the small values of λ implied by the estimated equations.

The second specification issue involves the treatment of episodes of government interference in the price-setting process. I introduce dummy variables for three such episodes: the NRA; price controls in World War II; and the Nixon era. Each dummy variable is constrained to sum to 1.0 over the years in which the program had its effect and to -1.0 over subsequent years when the program was dismantled. Thus each dummy variable sums to zero over its total period of impact, and its coefficient indicates the cumulative effect of the program in question.[10] We also allow for supply shocks by treating the relative price of food and energy as exogenous during 1947–78.[11]

The third specification issue involves the treatment of World War I, for which no dummy variables are created because there were no price controls. After Britain departed from the gold standard in 1914, agents rightly expected the structure of price setting to change from the pre-1914 norm. The significance of this temporary change in structure is tested by allowing coefficients to shift during the 1915–22 interval.[12] The burst of money creation in World War I presented a much greater contrast to the preceding gold standard era than did money creation in World War II (when the outbreak of war was preceded by four years of explosive growth in the monetary base during 1938–41). Another difference between the two wars was the widespread expectation in 1919–20 that the United Kingdom and United States would attempt to return to the prewar gold price, requiring the extinction of much of the fiat money created during the war, whereas the Fed's interest rate pegging policy prevented the development of any such expectation in 1946–47. Thus it is perhaps not surprising that, although structural coefficients were allowed to shift during World War II, no significant shift could be found other than the direct impact of the price controls.

The final specification issue involves the interaction of inflation expectations and special factors. Agents were smart enough to know that the end of wars in 1918 and 1945 made lagged prices an invalid predictor of future price behavior, and they presumably were also aware of the dismantling of NRA and the end of the Nixon era price controls. To reflect the assumption of intelligent expectation formation, the lagged price-change term used as one of the expectations proxies is constructed "net" of the estimated contribution of the dummy variables and the food-energy contribution (thus requiring iterative estimation). In addition, agents are assumed to have ignored the Great Depression and wars by setting the net lagged price-change term equal to zero in 1915–22 and 1929–49. While there is no important difference between the fit of the gross and net lagged inflation variables before 1953, the introduction of the net variable cuts in half the unexplained variance in the 1953–78 subperiod.

III. The Estimated Equation for 1892-1978 and Three Subperiods

Estimates of equation (6) are shown in Table 1. The first two lines exhibit the coefficients for the first two variables in (6). In the next two lines lagged net price change and nominal *GNP* change are proxies for expected inflation. Lagged changes in the money supply (M_2) were also introduced but appear mainly to be collinear with

[10]An additional dummy variable for The Korean War price controls proved to be insignificant. The precise values of the dummy variables are chosen to provide a measure of the timing impact of each program in annual data:

NRA Dummy		World War II Dummy		Nixon Era Dummy	
1933	0.4	1943	0.5	1972	0.5
1934	0.6	1944	0.4	1973	0.5
1935	−0.4	1945	0.1	1974	−0.3
1936	−0.6	1946	−0.6	1975	−0.7
		1947	−0.4		

[11]The "food-energy" variable in Table 1 is the difference for 1947–78 between the annual rates of change of the personal consumption deflator and the deflator net of expenditures on food and energy.

[12]The importance for expectation formation of special events, such as wars and postwar-adjustment periods, is stressed in my 1973 article.

TABLE 1—ESTIMATED EQUATIONS FOR THE ANNUAL PERCENTAGE CHANGE OF THE *GNP* DEFLATOR
(t-Ratios in Parentheses)

	1892–1978 (1)	1892–1929 (2)	1929–53 (3)	1953–78 (4)
Adjusted Nominal *GNP* Change ($\hat{y}_t$)				
a. Entire Period	.316	.287	.348	.341
	(8.83)	(3.34)	(9.65)	(8.47)
b. Extra effect, 1915–22	.306	.358	–	–
	(5.13)	(3.48)		
Lagged Output Ratio ($\hat{Q}_{t-1}$)				
a. Entire Period	.044	.176	.073	.130
	(1.86)	(1.63)	(3.61)	(3.57)
b. Extra effect, 1953–78	.090	–	–	–
	(0.60)			
Lagged "Net" Price Change (p^N_{t-1})				
a. Entire Period	−.225	−.041	−.731	.722
	(−1.49)	(−0.17)	(−2.96)	(10.09)
b. Extra effect, 1953–78	.836	–	–	–
	(2.80)			
Lagged Adj. Nom. *GNP* Change ($\hat{y}_{t-1}$)				
a. Entire Period	.179	.059	.147	–
	(3.99)	(0.57)	(3.21)	
b. Extra effect, 1915–22	.120	.209	–	–
	(1.86)	(1.82)		
c. Extra effect, 1953–78	−.305	–	–	–
	(−1.31)			
Special Factors				
a. NRA dummy, 1933–36	8.94	–	8.45	–
	(3.90)		(5.00)	
b. World War II dummy, 1943–47	−17.85	–	−17.39	–
	(−7.62)		(−10.54)	
c. Nixon Era dummy, 1972–75	−4.77	–	–	−4.70
	(−2.35)			(−9.88)
d. Relative Price of Food and Energy	.985	–	1.222	.493
	(2.17)		(2.58)	(3.56)
R^2/Standard Error	.877/2.10	.852/2.74	.961/1.41	.964/0.49
Durbin-Watson statistic	1.89	2.10	1.50	1.72

lagged nominal *GNP*.[13] The "special factor" variables are listed last. The four columns of the table display the results of estimation for the entire sample period and three subperiods.

The remarkable stability across subperiods of the coefficients on $\hat{y}_t$ is evident, as is the significantly greater response of prices to $\hat{y}_t$ during World War I. The effect of the Phillips curve variable, the lagged output ratio, is surprisingly small—real *GNP* slack of 5 percent (an output ratio of .95) slows inflation by only two-thirds of a percentage point per year in the postwar subperiod.

The shift in the role of lagged price change from regressive to extrapolative expectations stands out, both in the shift from negative coefficients to a positive coefficient in the last subperiod, and to the significance of the "extra 1953–78 effect" in the full sample period. This shift in structure was introduced in two alternative ways to compare Cagan's hypothesis that the recession flexibility of prices has gradually diminished during postwar recessions between 1949 and 1970, and an alternative hypothesis that the new mode of expectation formation

[13]Versions with lagged money perform better in the first subperiod, but substantially worse in 1929–53, reflecting the looseness of the money-*GNP* relationship during the Great Depression emphasized in my article with Wilcox.

TABLE 2—COMPARISON OF STANDARD ERRORS

	1892–1978	1892–1929	1929–53	1953–78
Standard Errors from:				
Table 1	2.10	2.74	1.41	0.49
Phillips Curve Alternative	3.83	4.70	4.48	0.79
Standard Deviation of Price Change	5.52	6.52	6.02	2.33

suddenly began in 1953. The Cagan hypothesis is tested in both columns (1) and (4) by multiplying the applicable lagged price-change variable by the time trend that moves smoothly from zero to unity between 1953 and 1970 and remains at unity thereafter. The Cagan variable does significantly worse in the postwar subperiod and cannot explain why inflation was so low during 1961–65 in the face of rapid nominal *GNP* growth.

How well do the equations in Table 1 fit as compared to the simple *EPC* alternative? Table 2 is a comparison of the standard errors of the equations in Table 1 with an *EPC* specification of Sachs which includes only the current output ratio and four lagged values of the dependent variable.[14]

Results similar to those in Table 1 have been obtained for the same specification applied to wage-change data. The similarity of the price and wage results suggests that the increased inertia observed in the postwar period characterizes wages and prices together rather than any shift in the cyclical behavior of the real wage. The annual change in the real wage responds negatively to the lagged output ratio, and positively to $\hat{y}_{t-1}$ (before 1953), to the NRA, and to both the World War II and Nixon era dummy variables.

IV. Implications

The results in Table 1 establish that the elasticity of price change to nominal *GNP* change has been approximately one-third for almost a century, and that the serial correlation in the price-change variable has shifted from negative to positive at some point in the early 1950's.[15] The inertia in the price-change process in the postwar period tends to dissuade policymakers from halting inflation, because a sluggish response of price change to restrictive demand policy creates high unemployment and political pressure to abandon the tight policy. But the extremely simple equation displayed in column (4) of Table 1 for the postwar years suggests a much greater payoff to restrictive demand policy than has recently been believed. An artificial (and implausible) experiment which drops adjusted nominal *GNP* growth from 6 to 0 percent causes inflation to slow down 1.8 percentage points in the first year, another 1.8 points in the second year, and 1.7 points in the third year, for a total response of 5.3 points after only three years.[16]

[14]To emphasize the contribution of the $\hat{y}$ variables in Table 1, the four "special factor" variables are introduced into the Sachs specification in addition to his dummy variable for World War I.

[15]Two questions of interpretation can be raised about the results in Table 1. First, the mere fact that the rate of price change depends on the change in $\hat{Q}$ does not by itself rule out a traditional Phillips curve interpretation, since Sachs has shown in his equation (10) that it is possible to combine my equation (1) above with an adaptive expectations specification of the determination of P_t^e and derive a reduced-form relationship between p_t, p_{t-1}, and $\hat{q}_t$. However this approach requires that the coefficient on p_{t-1} be constrained to equal 1.0, a constraint that seems quite inconsistent with the results in Table 1 for the pre-1953 period. Also, Sachs' estimates of his equation (10) for 1950–75 have a standard error three times higher than my postwar equation in Table 1. The second issue of interpretation concerns the low estimate of $\lambda = 0.4$ for 1953–78 implied by Table 1; this anomaly suggests to me that (2) and (3) may be too restrictive a structural specification and that the results in Table 1 may better be viewed as supporting a general structural dependence of price change on both the level and rate of change of output, as in (4).

[16]Of course the output ratio falls rapidly as well from 1.00 to .929 at the end of the third year. Then a subsequent policy in the fourth through sixth years of 3.0 percent adjusted nominal *GNP* growth will leave the economy with an inflation rate of only 1.3 percent and an output ratio of .993.

The change in the structure of expectation formation in the postwar period also reminds us that the conclusions of many econometric studies may be sensitive to extensions of sample period.[17] And it seems quite consistent with a change in attitude in the first postwar decade toward recognition of a fundamental change in the stabilizing role of government policy (initiatives based more on the automatic stabilizers and new institutions like FDIC than on countercyclical policy). The shift also emphasizes the crucial role of three-year staggered-wage contracts, a unique American institution that dates back to the first postwar decade and that introduces positive serial correlation in the wage-change data which in turn leads rational economic agents to expect positive serial correlation in the price-change data.

Some proponents of the classical equilibrium approach to macroeconomics, particularly Robert Barro, protest that, because these wage-setting institutions impose a high cost on some workers who experience employment fluctuations, they must not exist. Barro's position ignores the fact that rational firms and unions have chosen this bargaining pattern to minimize the real private costs of negotiations and strikes. The alternative classical equilibrium explanation of business cycles, that output responds positively to price "surprises," seems an implausible description of the last two decades in terms of the price equation developed in this paper, since the prolonged output boom of 1965–69 followed five straight years (1963–67) when price change was lower than the postwar equation can explain, while the "great recession" of 1974–75 occurred after six straight years (1970–75) when price change was *higher* than the fitted value.[18]

[17]The findings of Feige and Pearce that money contributes nothing to the explanation of prices, when the influence of lagged prices is held constant, is likely to be highly dependent on their 1953–71 sample period.

[18]The cumulative price error of −2.01 percentage points in 1963–67 can be contrasted with a swing in the output ratio ($\hat{Q}$) of +8.8 percentage points in 1961–66; the cumulative price error of +2.6 percentage points in 1970–75 can be contrasted with a swing in the output ratio of −7.8 percentage points in 1973–75.

REFERENCES

R. J. Barro, "The Equilibrium Approach to Business Cycles," work. paper, Univ. Rochester, Nov. 1979.

P. Cagan, "Changes in the Recession Behavior of Wholesale Prices in the 1920's and Post-World War II," *Explor. Econ. Res.*, Winter 1975, *2*, 54–104.

R. N. Cooper and R. Z. Lawrence, "The 1972–75 Commodity Boom," *Brookings Papers*, Washington 1975, *3*, 671–715.

E. L. Feige and D. K. Pearce, "Economically Rational Expectations: Are Innovations in the Rate of Inflation Independent of Innovations in Measures of Monetary and Fiscal Policy?," *J. Polit. Econ.*, June 1976, *84*, 499–522.

R. A. Gordon, "Wages, Prices, and Unemployment, 1900–1970," *Ind. Relat.*, Oct. 1975, *14*, 273–301.

R. J. Gordon, "Interest Rates and Prices in the Long Run: A Comment," *J. Money, Credit, Banking*, Feb. 1973, *5*, Part II, 460–63.

______, "Can the Inflation of the 1970s be Explained?," *Brookings Papers*, Washington 1977, *1*, 253–77.

______, *Macroeconomics*, Boston 1978.

______ **and J. A. Wilcox,** "Monetarist Interpretations of the Great Depression: An Evaluation and Critique," in Karl Brunner, ed., *Contemporary Views of the Great Depression*, Rochester 1980.

E. M. Gramlich, "Macro Policy Responses to Price Shocks," *Brookings Papers*, Washington 1979, *1*, 125–78.

B. Klein, "The Social Costs of the Recent Inflation: The Mirage of Steady 'Anticipated' Inflation," in Karl Brunner and Allan Meltzer, eds., *Institutional Arrangements and the Inflation Problem*, Vol. 3, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 185–212.

A. Meltzer, "Anticipated Inflation and Unanticipated Price Change," *J. Money, Credit, Banking*, May 1978, *8*, 182–205.

J. Sachs, "The Changing Cyclical Behavior of Wages and Prices: 1890–1976," *Amer. Econ. Rev.*, Mar. 1980, *70*, 78–90.

T. J. Sargent, "A Note on the 'Accelerationist' Controversy," *J. Money, Credit, Banking*, Aug. 1971, *3*, 721–25.

Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered

By CHRISTOPHER A. SIMS*

When monthly data on production, prices, and the money stock are interpreted via a vector autoregression, as generated by dynamic responses to "surprises" in each of the variables, a remarkable similarity in dynamics between interwar and postwar business cycles emerges. Nevertheless the size of the surprises is much larger in the interwar period. Furthermore, the money stock emerges as firmly causally prior, in C. W. J. Granger's sense, in both periods and accounts for a substantial fraction of variance in production in both periods.

When a short interest rate is added to the vector autoregression, the remarkable similarity in dynamics between periods persists, but the central role of the money stock surprises evaporates for the postwar period. While there are potential monetarist explanations for such an observation, none of them seem to fit comfortably the estimated dynamics. A nonmonetarist explanation of the dynamics, based on the role of expectations in investment behavior, seems to fit the estimated dynamics better. That this explanation, which is consistent with a passive role for money, could account for so much of the observed postwar relation between money stock and income may raise doubts about the monetarist interpretation even of the interwar data.

I. Monetarism and Evidence

I take monetarism to be the view that monetary policy is of central importance in the business cycle and that the time path of the money stock is a good single index of monetary policy. As set forth by Milton Friedman and Anna Schwartz, monetarism emphasizes the relation of the level of the money stock to the level of aggregate real economic activity, without a detailed theory of why money fails to be neutral in the short run. In its more recent guise, as surveyed recently by Robert Barro, monetarism develops an explicit basis for nonneutrality by positing barriers to information flow about prices.

Whether in its earlier or more recent form, monetarism claims support in the observed behavior of aggregate economic time-series. At least over some time periods, the money stock and income are highly correlated. Such correlation, while it is an implication of the theory and hence corroborates it, is easy to explain as noncausal, representing a passive response of the money stock to real activity. Friedman and Schwartz therefore have documented a tendency for movements in the money stock or its rate of change to precede movements in aggregate activity. This is a more complicated implication of the theory, and hence is stronger corroboration than the correlation by itself. It is also harder to explain as a passive response of the money stock to real activity. James Tobin, however, showed that such timing patterns could be explained by a model in which money played a passive role.

Friedman and Schwartz did not rely only on statistical timing relationships, however. Through detailed analysis of historical episodes, they attempted to document the existence of major swings in the money stock which not only preceded major swings in real activity, but were not themselves reflex responses to developments in real activity. In the postwar period, though, the relatively smooth behavior of the money stock, and the acceptance by the government of full-employment goals make isolation of convincingly "non-reflex" movements in the

*University of Minnesota and National Bureau of Economic Research. Thomas Doan executed the computations. Funds to support this research came from NSF grant SOC-7818042.

money stock very difficult. At the same time, the prewar episodes involve for the most part banking panics and international capital movements. The panics are almost inevitably sudden and unanticipated, but neither they nor the capital movements are ordinarily without antecedents in real economic activity. Furthermore, even if one accepts such episodes as shocks to the money stock which produced subsequent real developments, it is not obvious that one should extrapolate the dynamics of such events to the postwar period, where the movements in the money stock are thought to represent deliberate government policy moves to a much greater extent. Thus my 1972 demonstration that the money stock could be taken as exogenous in *GNP* on money-stock distributed-lag regressions was an important piece of support for the monetarist position. Despite the possibility that a substantial part of money-*GNP* correlation in the postwar period represented policy responses to developments in the economy, the data showed no evidence of such feedback; the observed statistical correlations and timing relationships were consistently interpretable as representing entirely causal effects of money on income.

Modern rational expectations monetarism has shifted attention away from structural interpretation of distributed-lag regressions of *GNP* on money stock. Nonetheless the fact that the money stock is causally prior to *GNP* in Granger's sense in postwar *U.S.* data is important for the modern monetarist position. Rational expectations monetarism suggests that it is surprises in movements in the money stock which generate nonneutrality. This implies a difference in the way data are examined for support for the monetarist position. Instead of finding the percentage of variation in real activity which can be explained by a distributed lag on the levels of the money stock, one looks for the percentage which can be explained by a distributed lag on surprises in the money stock. Now when "surprise" is taken to mean "innovation" in the technical time-series sense of "the prediction error in a best linear predictor," it is easy to show that Granger-causal priority of the money stock amounts to the equivalence of the percentages of variance in *GNP* accounted for by a distributed lag on the money stock and by a distributed lag on money stock surprises. Rational expectations monetarism yields a drastically different economic interpretation of the coefficients in distributed-lag regression of output on money, but it gives the same interpretation to the substantial fraction of variance explained by such regressions. With money Granger-causally prior, this fraction of variance represents, under new or old monetarist views, an unnecessary source of variability which could be eliminated by reform making monetary policy more predictable.

II. Innovation Accounting for Interwar and Postwar Data

A multivariate linear time-series model generates, according to the Wold decomposition theorem, a representation of each series in the model as a linear combination of current and past innovations in the variables in the system. These innovations are by construction serially uncorrelated, and if they are transformed to be contemporaneously uncorrelated as well, variance in the variables in the system can be unambiguously decomposed into components attributable to each innovation. The results reported in this paper come from autoregressive systems linear in the *logs* of the variables, using twelve lags of each variable, monthly data, and a constant term but no trend term. Estimation was by unconstrained least squares. The postwar period refers to 1948–78, using[1] data on 1947 for initial conditions, while the interwar period refers to 1920–41, using data on 1919 for initial conditions.

Table 1 shows that data on money, industrial production, and wholesale prices fit, in most respects, a familiar monetarist mold. For both periods, money is nearly entirely

[1]The methods are described in detail in my 1978 and 1980 papers. I intend that the results will be presented in more detail in a forthcoming discussion paper. Estimation was carried out with the assistance of Thomas Doan, using his recently minted program for econometric time-series analysis, (*RATS*).

TABLE 1—THREE-VARIABLE INNOVATION ACCOUNTING: PERCENTAGES OF 48-MONTH FORECAST-ERROR VARIANCE EXPLAINED (Interwar/Postwar)

Variables Explained	By Innovations in		
	*M*1	*IP*	*WPI*
*M*1	92/97	4/2	4/1
IP	66/37	28/44	6/18
WPI	38/14	19/7	43/80

Note: *M*1 = Money Stock; *IP* = Industrial Production; *WPI* = Wholesale Price Index.

accounted for by its own innovations, that is, it behaves as if it is Granger-causally prior. Tests of the hypothesis that all twelve lagged values of industrial production or of prices have zero coefficients in the money equation easily accept the null hypothesis. The smallest marginal significance level on these four *F*-tests is .18, confirming that the upper left corner of Table 1 is insignificantly different from 100 in both periods. Money innovations explain a substantial fraction of variance in industrial production in both periods, with the fraction notably more substantial in the interwar period. The fraction of price variance attributable to money innovations for the postwar period is smaller than what I had found in the earlier work with quarterly data already cited; this may be due at least in part to the use here of the more volatile *WPI* in place of the implicit price deflator, so that the long-run component price variance is a smaller portion of the total.

In both periods, the patterns of response of the system to innovations in the variables largely fit the monetarist framework. Production and prices respond positively to money innovations, both responses being smooth in both periods. Somewhat at variance with rational expectations monetarism is the lack of a tendency for production responses to money to be temporary in either period. Though both periods' responses peak at about 18 months, neither has decayed to half its peak level after four years. Despite the tendency of monetary shocks to persist in both periods, price responses in the interwar (not the postwar) period do show up as temporary, with the price response gone after four years. Production responses to a given shock in the *log* of money are larger in the postwar period, and price responses are smaller. This type of result has been interpreted in some recent work as evidence of greater price rigidity postwar, yielding greater real effects of given nominal surprises.

The most striking difference between the periods is in the variances of the innovations. Innovations in the *log* of money have a larger variance in the interwar period by a factor of about 22, for prices the factor is about 13.5, and for production the factor is 5. This fits the monetarist story that larger real fluctuations should be associated with larger monetary surprises, though the large difference in production innovation variances suggests that not all of the difference between periods is attributable to monetary policy and institutions—as most monetarists would certainly agree. Contemporaneous correlations among innovations are all much weaker in these monthly data than in quarterly data. For the postwar period they are not significantly different from zero; for the interwar period output innovations have significant correlations of .22 and .30, respectively, with money and prices.

Let us turn now to the more exotic pattern of results which emerges when short interest rates (the rate on 4–6 month prime commercial paper) are introduced into the system. I had found in earlier work with larger (nine variable) systems of quarterly data for the United States and Germany and of annual data for the United States that the proportion of variance in real variables attributable to money innovations shrank considerably in the larger systems. As Yash Pal Mehra's results would lead one to expect, Table 2 shows that with interest rates included, the money stock is no longer strongly Granger-causally prior. This result is in itself not counter to the monetarist position; the strikingly nonmonetarist aspect of Table 2 is that in the postwar period at the 48-month horizon only 4 percent of the variance of production is accounted for by money innovations. If this result is taken

TABLE 2—FOUR-VARIABLE INNOVATION ACCOUNTING PERCENTAGES OF 48-MONTH FORECAST-ERROR VARIANCE EXPLAINED (Interwar/Postwar)

Variables Explained	By Innovations in			
	R	*M*1	*WPI*	*IP*
R	63/50	28/19	7/4	1/28
*M*1	39/56	58/42	1/1	1/1
WPI	1/2	54/32	43/60	3/6
IP	16/30	58/4	7/14	18/52

Note: See Table 1. R = Short-Term Interest Rate.

at face value, a rational expectations monetarist must admit that surprise changes in the money stock have in fact played a trivial role in postwar business cycles; imposition of a monetarist rule to make the quantity of money more predictable would have had little real effect.

If one examines the moving-average representation (partially described in Table 3) in detail, one finds that the response of the *log* of production to a surprise unit increase in the *log* of the interest rate is essentially zero for about 6 months, followed by a smooth decline reaching a minimum around 18 months, with the minimum at $-.17$ with interwar data and at $-.23$ with the postwar data. After 48 months, the output response has in the interwar data begun to turn back down again, being by this point $-.20$, and in the postwar data, it has begun turning back up, being $-.12$. For the *log* of money stock, responses to an upward unit surprise in the *log* of interest rate are also in the form of a sustained, smooth decline. The shapes of these responses are similar across periods and their differences are marginally statistically significant at most, as can be seen from Table 3.

Thus in both periods some of the observed comovements of industrial production and money stock are attributed to common responses to surprise changes in the interest rate. With this shift in attribution, surprise changes in the money stock are left with a very small role in explaining production variance in the postwar period.

In nearly every case shown in Table 3, estimated response patterns are smooth in between the points for which data are displayed. While the responses are broadly similar, there are apparently important differences in the responses of interest rates to money and production; both these responses being much stronger in the postwar period. Also, response of production to prices is significantly negative in the postwar period in the first year, and is not negative in the first year in the interwar period. Because of the computational expense, standard errors have not yet been calculated for the interwar responses, so some of these apparently significant differences between the periods may not be in fact. A *chi*-square test for constancy of the dynamics, scaling residual variances in the triangularized autoregression to be constant across periods, yields a χ^2 (202) = 378.2. While this would certainly reject the null hypothesis of constancy if the asymptotic distribution theory were taken seriously, it is smaller than the Akaike criterion which aims at rejecting only restrictions "false enough" to increase mean square prediction error.

III. Possible Monetarist Explanations

A rational expectations monetarist, to avoid the conclusion that monetary policy surprises are not important in explaining the real component of postwar business cycles, must argue that in the results described above monetary policy surprises are being mismeasured. One possibility is that interest rate and monetary surprises are being confounded. The decompositions in Table 2 use a triangular orthogonalization of the innovations, in effect attributing forecast error variance to effects of interest innovations, and so on down the list displayed in the tables in the order interest, money, prices, production. This ordering was chosen because it maximizes the extent to which interperiod differences show up as differences in innovation variances, rather than differences in responses to innovations. However, because the postwar data yield such small correlations among innovations, the results that money innovations account for a trivial proportion of production variance

TABLE 3—RESPONSES TO UNIT SHOCKS

Variable shocked	Months later	Responses: Interwar, Postwar *R*		*M*1		*WPI*		*IP*		Approximate Postwar Standard Error *R*	*M*1	*WPI*	*IP*
R	1	1.0	1.0	−.05	−.01	.02	.01	−.04	.02	.02	–	.001	–
	3	1.1	1.4	−.06	−.03	.02	.01	.01	.05	.06	.004	.01	.02
	8	.89	.68	−.12	−.07	.05	.02	−.07	−.09	.11	.01	.02	.05
	16	.59	−.07	−.15	−.10	.02	.01	−.13	−.24	.23	.02	.04	.07
	24	.65	−.59	−.19	−.10	−.01	−.01	−.14	−.19	.27	.02	.05	.07
	48	.46	−.11	−.23	−.10	.01	−.04	−.20	−.12	.27	.02	.04	.06
*M*1	1	0.0	0.0	1.0	1.0	.01	.10	.47	.42	–	–	.03	.06
	3	−.25	3.02	.94	1.19	.34	.24	.85	.95	1.40	.08	.15	.26
	8	.06	10.08	1.32	1.52	1.18	.71	2.43	1.51	2.78	.26	.33	.57
	16	2.32	11.71	1.76	1.51	1.80	1.40	3.06	1.58	4.62	.36	.64	1.20
	24	3.25	13.38	1.63	1.34	1.31	1.76	1.89	.16	4.92	.38	.87	1.04
	48	3.56	−.31	.51	1.37	−.06	2.04	.51	.53	2.61	.48	1.03	1.04
WPI	1	0.0	0.0	0.0	0.0	1.0	1.0	.39	.03	–	–	–	–
	3	−.31	.79	.17	.03	1.39	1.37	1.15	−.12	1.0	.05	.07	.20
	8	.93	−1.84	.25	−.08	1.58	1.74	.58	−.78	1.9	.11	.22	.43
	16	1.21	−.08	.08	−.03	.67	1.70	−.46	−1.51	2.6	.20	.39	.54
	24	1.14	−1.26	−.06	−.06	.26	1.38	−.66	−1.53	2.0	.28	.47	.72
	48	1.32	−2.18	−.21	−.22	.16	.78	−.10	−.81	2.1	.43	.47	1.02
IP	1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	–	–	–	–
	3	−.33	1.37	.09	.05	.05	.12	1.48	1.48	.42	.02	.04	.11
	8	.65	3.40	.08	.03	.17	.24	.94	1.24	.98	.05	.08	.21
	16	.34	3.24	.21	−.01	.28	.25	.48	.72	.97	.08	.14	.27
	24	.22	1.32	.18	0.0	.13	.05	.11	.49	.97	.10	.16	.32
	48	.24	1.45	.01	.09	−.06	−.23	.02	.87	1.03	.15	.24	.47

is robust to the ordering of the orthogonalization. In other words, there is not much relation in the data between interest and money surprises.

But the innovations might be mismeasured because the time unit is wrong. If the time delay relevant to rational expectations business cycle theory is longer than a month, it may be that some of the true money-stock surprise shows up spuriously as interest-rate surprise with this fine time unit. This possibility seems ruled out, however, by the fact that the decompositions of variance with annual data show precisely the same antimonetarist phenomenon—money surprises account for less than 10 percent of output variance when an interest rate is included in the system.

What about the possibility that some people in fact often anticipate policy-induced movements in the money stock? In this case, one might expect the interest rate to rise in anticipation of forthcoming monetary tightness. If in addition the true time delay relevant to the rational expectations theory exceeds a month, one might then get the pattern of results we have displayed. This line of argument deserves further exploration, but it is not immediately clear that it can avoid internal contradictions. It certainly requires that some economic agents ignore published information on current interest rates.

A monetarist not maintaining the rational expectations stance might have an easier time explaining the results. If one is not claiming that changes in the money stock must be unanticipated in order to have a real effect, the notion that some money-stock changes are anticipated, and therefore preceded by upward movements in short interest rates, is quite acceptable. In fact one reason that this might happen leaps to mind.[2] Changes in base money might be transmitted to the stock of currency and demand deposits only with a delay, while having quick effects on the interest rate.

When the postwar system is estimated with reserves or base money replacing the

[2]It leaped to my mind, however, only after Robert Gordon had pointed it out to me.

money stock, however, almost precisely the same pattern of results emerges. The percentages of variance in industrial production explained by money innovations remain at or below 10 percent. The only notable difference is that base money, unlike currency plus demand deposits or reserves, shows no negative response to interest-rate innovations; production still shows the same negative response to interest-rate innovations in the systems with base money.

More generally, there is another difficulty with interpreting interest-rate innovations as simply anticipated movements in the money stock. For both interwar and postwar data, the price level responds to money shocks with a steady price inflation over a year, while interest-rate shocks, despite their effects on money supply, produce no substantial effect on prices. If interest-rate innovations are simply anticipated money-stock innovations, it is hard to see why they should affect prices so differently. Of course the rational expectations monetarist view does predict a difference here, but of the opposite sort—anticipated money-stock changes should have more effect on prices.

In the interwar years there were "panics" and in the postwar years there were "liquidity crunches." If these are interpreted as shifts in the public's preferences toward cash, away from deposits, they might be the source of the observed response to interest innovations. If, as the public tries to convert deposits to cash, the Federal Reserve responds weakly or not at all with injections of reserves, one would expect a quick rise in interest rates, a fall in the money stock, and a decline in output as if there had been a deliberate monetary tightening. This story is not "monetarist" in the sense I gave the term at the outset, in that it does not attribute the observed pattern to surprises in monetary policy directly. On the other hand, this story is in the spirit of Friedman and Schwartz's own discussion of the depression, in which they claim not that the initial shocks came from arbitrary monetary policy, but rather that failure of monetary policy to respond appropriately to shocks originating elsewhere magnified the effects of those shocks.

This explanation is not implausible to me. It does have defects. It leaves open the question of why price responses to this type of shock are different from those to innovations in money supply. It seems to require that the monetary authorities in the postwar period respond in almost the same pattern to an increased demand for cash as did the monetary authorities in the interwar period, which might seem implausible. And it leaves unexplained the origin of these sudden, cyclically important shifts in the demand for cash.

IV. A Nonmonetarist Expectational Theory to Fit the Facts

A Keynesian view of the business cycle centers attention on the relation of capital purchases to expectations of future profitability. As is now widely understood, in order for expectations of the future to play the central role in investment behavior which Keynesian theory gives them, it must be costly to adjust the capital stock rapidly. The theory which emerges is much the same, whether one has adjustment costs internal to the firm or external, in the form of a capital goods industry with increasing costs. In the latter case, firms which are capital goods price takers will have as an equilibrium condition

$$(1) \qquad r = DP_k/P_k + \pi/P_k$$

where P_k is the effective price of capital goods (including discounts, the cost of obtaining prompt delivery, etc.), r is the instantaneous interest rate, and π is the real marginal product of a physical unit of capital. D indicates differentiation with respect to time. Suppose information becomes available indicating that the real yield on capital, π, will decline at some point several months from now. It seems plausible that this would lead to a drop in the rate of investment, and hence to a drop in P_k. If this drop in investment is persistent over several months, DP_k must remain small initially. From (1) above we can see that this means that r must rise.

This story does of course depend on some implicit assumptions. If P_k is held rigid either by a very flat capital goods supply

curve, or by a rate of saving which is insensitive to returns, even over the short run, then (1) will be satisfied by a persistently tight link between r and π. Knowledge of a future decline in π could not then raise current r.

Clearly this story fits the response of production to interest-rate innovations, in particular to the 6-month period following the shock, in both interwar and postwar response patterns, during which production remains flat. The observed responses of money stock to the interest shocks could simply be the tail following the dog: nonmonetary economic developments raise interest rates, then push production down; and the demand for money declines smoothly in response, as standard theories lead one to expect.

This theory explains the similarity in response to interest shocks across periods by similarity in the short-run supply elasticity for capital goods and similarity in short-run yield elasticities of savings. This seems more plausible to me than the similarity of persistent patterns of monetary policy errors which the monetary theories seem to require. The theory does not directly explain why price responses to interest and to money-stock innovations should be different, but such differences are certainly no paradox from the point of view of the theory. For monetarist theories, the absence of price response to a change in money stock following an interest-rate surprise does seem a problem.

It should be noted that this theory is not contradictory to the interpretation of interest-rate shocks as representing liquidity crunches. The interest-rate surprise in this theory represents a surprise decline in valuation of existing assets while current real productivities of capital remain high. One would expect such a situation to result in problems in maintaining collateral for bank loans and complaints that loans for legitimate working capital purposes are available only at high interest rates.

V. Conclusions and Implications

The theory in the preceding section has no direct implications for whether active countercyclical monetary or fiscal policy can have good effects, or any effects. Even as a working hypothesis, however, the theory raises some interesting issues. It treats an historically reliable pattern of dynamic statistical relations, which look like causal relations ought to look, as reflective of the workings of anticipations through financial markets. It has long been recognized (as pointed out in some detail in my 1977 paper) that prices of freely traded durable goods, especially including financial assets, should behave to a close approximation as if "Granger-causally prior" to any time-series observable by market participants. The stock of money is not the price of an asset, and we are used to thinking of it as determined by the Federal Reserve, with shifts in demand for money having little immediate impact on the stock. But the demand for money ought certainly in principle to be related to the value of existing assets. If we view the stock of money as quickly responsive on a month-to-month basis to shifts in demand for it, the prospect arises that distributed-lag regressions of production on money have predictive value for the same reason that similar regressions using stock prices do. A theory which rigorously developed this possibility would amount to a stochastic version of James Tobin's "Money and Income: Post Hoc Ergo Propter Hoc?" Exploring the implications of theory in this line seems to me a major item on the agenda for macro-economic research. Money innovations after all still seem to explain most of the interwar business cycle. Is this because surprises in monetary policy were really more important in that period, or would the result evaporate in a model which treated monetary surprises symmetrically with a wider array of financial surprises?

Appendix: Notes on the Tables

A linear model for a vector stochastic process x can be expressed as

$$x_t = \sum_{s=0}^{\infty} A_s e_{t-s}$$

where $e_t = x_t - E(x_t | x_{t-1}, x_{t-2}, \ldots)$. If we

then choose a lower triangular matrix B such that Be_t has a diagonal covariance matrix and B has ones on its diagonal, we can replace A by $C=AB^{-1}$ and e by $f=Be$, to obtain

$$x_t=\sum_{s=0} C_s f_{t-s}$$

For the linear model fit to *logs* of the variables of this paper, the coefficients in C are what is reported in Table 3 as "responses to innovations." The variance-covariance matrix of $x_t-E(x_t|x_{t-k}, x_{t-k-1}, \ldots)$, the k-period-ahead forecast of x, is given by

$$V_k=\sum_{s=0}^{k} C_s Var(f_t) C_s'$$

This formula, with $k=48$, is used to generate Tables 1 and 2. The approximate standard errors in Table 3 were generated by Monte Carlo integration of the likelihood, and correspond to the standard errors of Bayesian posterior distributions with a flat prior. They are approximate not mainly because of their Monte Carlo source, but rather mainly because they were generated with the data orthogonalized in a different order than that used to generate the responses tabulated. Because of the near-orthogonality of the postwar residuals, this makes little difference to the responses, but it does affect the standard errors of first and second period responses quite a bit, in percentage terms.

REFERENCES

H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 1974, AC-19, 716–23.

R. Barro, "The Equilibrium Approach to Business Cycles," work paper, Univ. of Rochester, Nov. 1979.

M. Friedman and A. J. Schwartz, "Money and Business Cycles," *Rev. Econ. Statist.*, Feb. 1963, *45*, Part 2, 32–64.

______ **and** ______, *A Monetary History of the United States*, Princeton, 1963.

C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, July 1969, *37*, 424–38.

C. A. Sims, "Money, Income, and Causality," *Amer. Econ. Rev.*, Sept. 1972, *62*, 540–52.

______, "Exogeneity and Causal Ordering in Macroeconomic Models," in *New Methods in Business Cycle Research*, Fed. Reserve Bank Minneapolis, 1977, 23–44.

______, "Small Econometric Models of the U.S. and Germany," disc. paper, Univ. Minnesota, Dec. 1978.

______, "Macroeconomics and Reality," *Econometrica*, Jan. 1980, *48*, 1–48.

J. Tobin, "Money and Income: Post Hoc Ergo Propter Hoc?," *Quart. J. Econ.*, May 1970, *84*, 310–17.

DISCUSSION

ALLAN H. MELTZER, Carnegie-Mellon University: Robert Gordon has undertaken to explain the rate of price change in each of the past eighty-seven years. I applaud his effort because I believe that much more can be learned from studies of annual data, despite the imprecision in earlier years, than from additional studies of quarterly data for the recent past.

During the period Gordon considers, the United States shifted from gold to a paper standard, and from fixed to floating exchange rates. The role of government in the economy expanded greatly. From studies of this experience, we can learn about the consequences of these and other changes for inflation, price, and output variability, and the formation of anticipations. Eventually, we will learn about the institutions that increase or reduce variability.

Gordon's principal findings are: 1) there is a nearly constant response of the percentage rate of price change to the percentage rate of change of nominal *GNP*; 2) the size of the response is approximately one-third; 3) there is a slower response by the prices observed during postwar recessions (mainly the result of changes in beliefs about persistent inflation and not, as commonly alleged, the result of increased powers of unions and monopolies or other changes in the relation between prices and output); 4) the standard Phillips curve is far less reliable than a relation that links output or spending to the price level so that the rate of price change varies with the rate of growth of output or spending, and not with the level of spending or output; and 5) the lagged inflation rate is now a far more important predictor of current inflation than before World War II.

Conclusions 3) and 4) are consistent with and supportive of my own earlier work, especially the rejection of the standard Phillips curve on the grounds that it is misspecified. Conclusion 5) is similar to my finding that the maintained average rate of money growth became much more important than the current rate of money growth once the United States left the gold standard. The greater importance of past average rates of change of money and prices does not mean that people now look farther back when forming anticipations about the future rate of inflation. The opposite is more likely to be true.

To see why, think of the current rate of price change p as consisting of two components, $p = \pi + \rho$. The anticipated rate of inflation π is the central value around which prices are expected to change; ρ is the one-time rate of change of prices. (See Gordon's equation (3).) Under the gold standard, π changed slowly, but ρ changed frequently. Consequently, people looked back farther to estimate π, but gave less weight to π when forming anticipations about p. In the 1950's and 1960's, π dominated p, so past average money growth (or past rates of price change) received greater weight in the expected rate of inflation. The oil shocks of the 1970's shifted some of the weight from π back to ρ.

Gordon's conclusions 1) and 2) differ from Phillip Cagan's and my own earlier result. The difference in conclusion is related to the way in which Gordon models the pricing process. His model, like many standard Phillips curves, is based on an implausible idea. Deviations of prices from expected price levels depend on the gap between current output and some measure of capacity. The problem does not depend on either the use of price levels instead of rates of price change or on the use of output instead of employment. Expected price levels are not related to expected levels of output except in the special case in which the economy is near capacity. The use of capacity involves sizeable error in the depressions of the 1890's and the 1930's.

Gordon overcomes the problem by introducing lagged "net" price change as a measure of anticipated inflation. Net price change is not significant in the heyday of the gold standard. The lack of significance warns us, and should have warned Gordon, that his explanation of a shift from regres-

sive to extrapolative price expectations is not correct. There was no reason under the gold standard to anticipate that prices would fall next year if they rose this year. That result depends not only on speed of adjustment, but also on whether the aggregate world stock of gold increased or was redistributed.

Gordon has produced some interesting results. The generous use of dummy variables and extra effects, the absence of a model of demand and the presence of current nominal *GNP* growth as a variable in an equation explaining the rate of price change warns us not to place much reliance on R^2, predictions, and precise numerical results.

RECENT DEVELOPMENTS IN THE ECONOMIC THEORY OF INDEX NUMBERS

Capital and the Theory of Productivity Measurement

By W. E. DIEWERT*

The purpose of this paper is twofold: to provide a brief survey of recent approaches to the measurement of total factor productivity; and to comment on some of the special problems which are associated with capital. Thus in Sections I–IV below, four approaches to the measurement of total factor productivity are outlined, while Section V compares and contrasts the approaches. In Section VI, I consider how the various approaches have to be modified when well-defined rental markets for components of the capital stock do not exist (the usual case).

I. Econometric Approaches to the Measurement of Total Factor Productivity

Let y^t be the output produced by a firm during period t and $x^t \equiv (x_1^t, x_2^t, \ldots, x_N^t)$ be the vector of inputs utilized during period t. Suppose that the firm's technology can be represented by a production function f_t in period t. I shall follow Robert Solow and identify changes in total factor productivity with shifts in the production function. In the typical econometric approach to measuring shifts in the production function, one assumes that the period-specific production functions $f_t(x)$ can be rewritten as $f(x,t)$. Then a convenient functional form for the production function f is assumed, and the unknown parameters which characterize f are estimated using the regression equation $y^t = f(x^t, t) + \text{error}$ for $t = 1, 2, \ldots, T$. Note that T is the number of periods for which data on output and inputs are available.

This model can be extended to the multiple output case: simply include any additional outputs in the input vector. However, in the multiple output case, there are econometric advantages to defining the joint cost function C as follows:

$$(1) \quad C(y,w,t) \equiv \min_x \{ w \cdot x : (y,x) \in S^t \}$$

where $y \equiv (y_1, y_2, \ldots, y_M)$ is a vector of outputs (if $y_m < 0$, then the mth "output" is an intermediate input), $w \equiv (w_1, w_2, \ldots, w_N) \gg 0_N$ is a positive vector of input (rental) prices that the producer faces, $x \equiv (x_1, x_2, \ldots, x_N) \geqslant 0_N$ is a nonnegative vector of inputs utilized, $w \cdot x \equiv \sum w_n x_n$, and S^t denotes the firm's period t production possibilities set. (If there is no x such that $(y,x) \in S^t$, then $C(y,w,t) \equiv +\infty$).

Obviously, the joint cost function $C(y,w,t)$ is completely determined by the production-possibilities set S^t. Under certain regularity conditions (for example, see W. M. Gorman, Daniel McFadden, or my 1974 paper), C determines S^t. Moreover, if C is differentiable with respect to the components of the input price vector w, we have the following useful result (Hotelling's or Shephard's Lemma):

$$(2) \qquad x^t = \nabla_w C(y^t, w^t, t)$$

where x^t, y^t, and w^t denote the period t input, output, and input price vectors, respectively, and $\nabla_w C$ denotes the vector of partial derivatives of C with respect to the components of w.

The second econometric approach to measuring shifts in the production function or the production-possibilities set consists

*University of British Columbia.

of: (i) assuming that the producer competitively minimizes costs; (ii) assuming a convenient functional form for the joint cost function C; (iii) adding errors to the system of input demand functions (2); and (iv) using the resulting system of equations in order to estimate econometrically the unknown parameters of C. Once C has been determined, summary measures of the shift in technology (such as $\tau(y^t, w^t, t) \equiv \partial \ln C(y^t, w^t, t)/\partial t$) can readily be calculated.

There are two important considerations in the choice of a functional form for C: the functional form should be capable of providing a second-order approximation to an arbitrary twice-continuously differentiable function; and the derivatives (or the logarithmic derivatives) of C should be linear in the unknown parameters in order to facilitate econometric estimation. The translog functional form, due to Laurits Christensen, Dale Jorgenson, and Laurence Lau, and to Dennis Sargan (see also my 1974 paper, p. 139), satisfies the above criteria. Empirical applications of the above cost function approach can be found in Hans Binswanger and in Ernst Berndt and David Wood (for the single output case) and in Randall Brown, Douglas Caves, and Christensen (for the multiple output case).

The two econometric approaches to measuring shifts in technology outlined above are consistent with increasing returns to scale and possible monopolistic behavior on the part of the firm on output markets. However, if we are willing to assume competitive profit-maximizing behavior on output markets as well as input markets, then the period t observed vector of outputs y^t will be a solution to the profit-maximization problem $\max_y \{p^t \cdot y - C(y, w^t, t) : y \in \tilde{S}^t\}$ where $p^t \equiv (p^t, p_2^t, \ldots, p_M^t) \gg 0_M$ is the vector of positive output prices that the firm faces during period t and $\tilde{S}^t \equiv \{y : \text{there exists } x \text{ such that } (y, x) \in S^t\}$ is the feasible output set. If C is differentiable with respect to the components of y and if y^t belongs to the interior of $\tilde{S}^t$, then the first-order conditions for the profit-maximization problem yield

$$(3) \qquad p^t = \nabla_y C(y^t, w^t, t)$$

Errors can be appended to equation (3), and the stochastic versions of (2) plus (3) can be used in order to estimate the unknown parameters occurring in the joint cost function C. Of course equations (3) are not consistent with increasing returns to scale, but they are consistent with the technology set S^t exhibiting constant or decreasing returns to scale.

II. The Divisia Approach

In this approach, it is no longer assumed that data are available only at discrete points in time; instead assume that data are available at every moment of time t where t ranges over a closed interval. Let us consider the single output case first.

Let $y(t) \equiv f(x(t), t)$ be output at time t and let $x(t) \equiv (x_1(t), x_2(t), \ldots, x_N(t))$ denote the vector of inputs utilized at time t. Following Solow and assuming that the production function f is differentiable, we differentiate the equation $y(t) = f(x(t), t)$ with respect to t and divide both sides of the resulting identity by $f(x(t), t)$. If we denote time derivatives with a prefix D, we obtain the identity $Dy(t)/y(t) = \sum_{n=1}^N [\partial \ln f(x(t), t)/\partial x_n] Dx_n(t) + \partial \ln f(x(t), t)/\partial t$. If the price of output is $p(t)$ at time t and the producer pays each input the value of their marginal product so that the nth input price $w_n(t) = p(t) \partial f(x(t), t)/\partial x_n$, then the above identity can be rearranged to yield

$$(4) \qquad \frac{\partial \ln}{\partial t} f(x(t), t) = \frac{Dy(t)}{y(t)} - \sum_{n=1}^{N} s_n(t) \frac{Dx_n(t)}{x_n(t)}$$

where $s_n(t) \equiv w_n(t) x_n(t)/p(t) y(t)$. Note that the left-hand side of (4) gives the period t proportional rate of growth in output which is unexplained by the growth in input. It is a convenient summary measure of technical progress at time t. If continuous data on output, inputs and output, and input prices were available, then the right-hand side of (4) could be calculated and thus the left-hand side of (4) could be determined. In empirical applications, the time derivatives

on the right-hand side of (4) are approximated by discrete differences (for example, see Solow or Jorgenson and Zvi Griliches).

Following Caves, Christensen, and Joseph Swanson, and Michael Denny, Melvyn Fuss, and Leonard Waverman, Solow's approach can be extended to the multiple output case: simply assume that the joint cost function C defined by (1) is defined and differentiable for each moment of time t and differentiate the identity $\Sigma_{n=1}^{N} w_n(t)x_n(t) = C(y(t), w(t), t)$ with respect to time t. Assuming competitive profit-maximizing behavior so that (2) and (3) hold for each instant of time t, and dividing both sides of the identity by $C(y(t), w(t), t)$, we obtain the following expression for $\tau(y(t), w(t), t) \equiv \partial \ln C(y(t), w(t), t)\partial t$:

$$(5) \quad \tau(y(t), w(t), t) = \sum_{n=1}^{N} \frac{w_n(t)x_n(t)}{w(t)\cdot x(t)} \frac{Dx_n(t)}{x_n(t)} - \sum_{m=1}^{M} \frac{p_m(t)y_m(t)}{w(t)\cdot x(t)} \frac{Dy_m(t)}{y_m(t)}$$

where $w(t)\cdot x(t) \equiv \Sigma_{n=1}^{N} w_n(t)x_n(t) = C(y(t), w(t), t)$. If there is no technological regress in the economy, then we have $S^r \subset S^t$ if $r < t$ (the production-possibilities sets do not become smaller as time marches on). Thus the partial derivative of $C(y(t), w(t), t)$ with respect to its last argument will not be positive and hence $\tau(y(t), w(t), t) \leqslant 0$ in the no-technological regress case. Note that $\tau(y(t), w(t), t)$ is the percentage rate of cost increase that is unexplained by changes in inputs and outputs of time t.

In empirical applications, the time derivatives on the right-hand side of (5) are approximated by discrete differences or an index number formula. For empirical applications plus additional material on this approach to the measurement of total factor productivity, see Denny, Fuss, and Waverman; Christensen, Dianne Cummings, and Jorgenson; and Frank Gollop and Jorgenson.

III. The Exact Index Number Approach

In the econometric approach to measuring shifts in total factor productivity, we generally have to approximate the firm's true cost or production function with a particular functional form. In the Divisia approach, we have to approximate continuous time derivatives with discrete differences. The exact index number approach also involves an approximation: let us assume that the producer's production function or joint cost function is a certain functional form, and then attempt to derive an index number formula that is consistent with the assumed functional form.

For example, assume that the firm's joint cost function is the following translog variable cost function:

$$(6) \quad \ln C(y^t, w^t, t) \equiv \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln y_m^t + \tfrac{1}{2} \sum_{m=1}^{M} \sum_{k=1}^{M} \alpha_{mk} \ln y_m^t \ln y_k^t + \sum_{n=1}^{N} \beta_n \ln w_n^t + \tfrac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{N} \beta_{ni} \ln w_n^t \ln w_i^t + \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma_{mn} \ln y_m^t \ln w_n^t + \sum_{m=1}^{M} \delta_m t \ln y_m^t + \sum_{n=1}^{N} \varepsilon_n t \ln w_n^t + \phi t + \chi t^2$$

where $\Sigma_{n=1}^{N} \beta_n = 1, \Sigma_{i=1}^{N} \beta_{ni} = 0$ for $n = 1, 2, \ldots, N$, $\Sigma_{n=1}^{N} \varepsilon_n = 0$, $\beta_{ni} = \beta_{in}$, $\alpha_{mk} = \alpha_{km}$, $y^t \gg 0_M$ and $w^t \gg 0_N$. Now if a function $g(z)$ is quadratic in a vector of variables z, then it can be verified that the following identity is true:

$$(7) \quad g(z^1) - g(z^0) = \tfrac{1}{2}\left[\nabla_z g(z^1) + \nabla_z g(z^0)\right] \cdot \left[z^1 - z^0\right]$$

Upon noting that $\ln C$ defined by (6) is quadratic in the logarithms of output quantities and input prices, $\ln y_m^t$ and $\ln w_n^t$, and

time t, we may apply the identity (7). If we also assume competitive profit-maximizing behavior so that (2) and (3) are valid at time periods t=0 and t=1, then the resulting identity can be rearranged to yield

$$(8)\ \tfrac{1}{2}[\tau^1+\tau^0]$$
$$=\left\{ ln\frac{w^1\cdot x^1}{w^0\cdot x^0}-\frac{1}{2}\sum_{n=1}^{N}\left[\frac{w_n^1x_n^1}{w^1\cdot x^1}+\frac{w_n^0x_n^0}{w^0\cdot x^0}\right]ln\frac{w_n^1}{w_n^0}\right\}$$
$$-\left\{\frac{1}{2}\sum_{m=1}^{M}\left[\frac{p_m^1y_m^1}{w^1\cdot x^1}+\frac{p_m^0y_m^0}{w^1\cdot x^1}\right]ln\frac{y_m^1}{y_m^0}\right\}$$

where $\tau^t \equiv \partial\, ln\, C(y^t,w^t,t)/\partial t$ for t=0, 1 is the period t impact effect on cost due to technological change. Note that the right-hand side of (8) can be calculated provided that output and input data on prices and quantities for periods 0 and 1 are available. Note further that the first term in curly brackets on the right-hand side of (8) is the logarithm of an (implicit) quantity index of inputs (the implicit Törnqvist quantity index $\tilde{Q}_0(w^0,w^1,x^0,x^1)$ which is discussed in my 1976 paper (p. 121) while the second term is the logarithm of a quantity index in outputs. If the technology exhibits constant returns to scale, then $p^t\cdot y^t=w^t\cdot x^t=C(y^t,w^t,t)$ for t=0, 1, and the quantity index in output will reduce to the Törnqvist quantity index $Q_0(p^0,p^1,y^0,y^1)$ discussed in my 1976 paper (p. 120). However, in general (8) is valid for a nonconstant returns-to-scale technology. Finally, we note that the translog joint cost function defined by (6) is capable of approximating an arbitrary differentiable joint cost function to the second order. Thus the measure of the shift in technology defined by the right-hand side of (8) should be fairly reliable provided that the producer is competitively maximizing profits.

Further examples of the exact index number approach can be found in Robert Pollak; Sidney Afriat (1972b); Paul Samuelson and S. Swamy; my 1976, 1977, 1978, 1979 papers; Christensen, Cummings, and Jorgenson; Gollop and Jorgenson; and Lau. Closely related material involving approximating preferences quadratically may be found in Henri Theil (1967, pp. 200–12; 1968) and T. Kloek.

IV. The Nonparametric Approach

Suppose we have a time-series of data on outputs and inputs $\{(y^t,x^t): t=1,2,\ldots,T\}$. If the technology sets S^t are concave and subject to free disposal and if the no-technological-regress assumption is satisfied so that $S^r\subset S^t$ if $r<t$, then it can be shown that the set $\{(y,x): y\leqslant\Sigma_{i=1}^t\lambda_iy^i, x\geqslant\Sigma_{i=1}^t\lambda_ix^i, \lambda_i\geqslant 0, \Sigma_{i=1}^t\lambda_i=1\}\equiv\tilde{S}^t$ forms an inner approximation to the true production-possibilities set S^t. In the single output case, following Michael Farrell, Afriat (1972a), and Giora Hanoch and Michael Rothschild, define the inner approximation period t production function $\tilde{f}^t$ over the set of input vectors $\{x: x\geqslant\Sigma_{i=1}^t\lambda_ix^i, \lambda_i\geqslant 0, \Sigma_{i=1}^t\lambda_i=1\}\equiv S(x^1,x^2,\ldots,x^t)$ by

$$(9)\quad \tilde{f}^t(x)\equiv\max_{\lambda_1}\geqslant 0,\ldots,\lambda_t\geqslant 0$$
$$\left\{\sum_{i=1}^t\lambda_iy^i: \sum_{i=1}^t\lambda_ix^i\leqslant x, \sum_{i=1}^t\lambda_i=1\right\}$$

for $t=1,2,\ldots,T$. (If the true production function exhibits constant returns to scale, then the constraint $\Sigma_{i=1}^t\lambda_i=1$ can be dropped from the linear programming problem (9) and the period t approximating production function $\tilde{f}^t$ will be well defined for all input vectors $x\geqslant 0_N$.)

The following (nonstatistical) test allows us to determine whether the given firm data are consistent with increasing total factor productivity over time.

Test 1 (see my paper with Celik Parkan): If $y^t=\tilde{f}^t(x^t)$ for $t=1,2,\ldots,T$, then the given data $\{(x^t,y^t): t=1,2,\ldots,T\}$ are consistent with the efficiency hypothesis (i.e., the firm produces a maximal amount of output for a given input vector) and with the hypothesis of no technological regress for some family of production functions $\{f^t: t=1,2,\ldots,T\}$ where f^t is continuous, nondecreasing, and concave over the domain set $S(x^1,x^2,\ldots,x^t)$. On the other hand, if $y^t<\tilde{f}^t(x^t)$ for any t, then the given data are not consistent for any family of continuous, nondecreasing, and concave production functions $\{f^t\}$ satisfying the no-technological-regress assumption.

If the given data satisfy Test 1, then a reasonable measure of the shift in the true production function going from period $t-1$ to period t is $\tilde{f}^t(x^{t-1})/\tilde{f}^{t-1}(x^{t-1})$.

The above approach does not make use of the assumption that the producer is minimizing costs, an assumption that I now make. Suppose that data on input prices $w^t \gg 0_N$, inputs $x^t \geqslant 0_N$, and outputs $y^t \geqslant 0_M$ are given for T time periods and that S^t denotes the period t production-possibilities set. The period t input production-possibilities set for a given output vector $y \geqslant 0_M$ L^t is defined as $L^t(y) \equiv \{x:(y,x) \in S^t\}$. Also define the set of indices M_t for $t=1,2,\ldots,T$ by $M_t \equiv \{i: i$ is an integer, $1 \leqslant i \leqslant t$ and i is such that $y^i \geqslant y^t\}$.

Note that $t \in M_t$ for all t. Now we can define the following linear programming problems for $t=1,2,\ldots,T$:

(10)

$$\min_{\lambda_i}\left\{\sum_{i \in M_t} \lambda_i w^t \cdot x^i : \lambda_i \geqslant 0, \sum_{i \in M_t} \lambda_i = 1\right\} \equiv C^{t*}$$

Note that $C^{t*} \leqslant w^t \cdot x^t$ since $t \in M_t$. The following test allows for multiple outputs.

Test 2 (see my 1979 paper): If $C^{t*} = w^t \cdot x^t$ for $t=1,2,\ldots,T$, the given data $\{(w^t, x^t, y^t): t=1,2,\ldots,T\}$ are consistent with cost-minimizing behavior for some family of production-possibilities set S^t which satisfies the no-technological-regress assumption, where each S^t is a closed nonempty set which has the free disposal property and is such that $L^t(y^t)$ is a convex set (call these *Conditions I* on $\{S^t: t=1,\ldots,T\}$). In this case, we can define an inner approximation to the set $L^t(y^t)$ by

(11)

$$\tilde{L}^t(y^t) \equiv \left\{x: x \geqslant \sum_{i \in M_t} \lambda_i x_i, \lambda_i \geqslant 0, \sum_{i \in M_t} \lambda_i = 1\right\}$$

On the other hand, if $C^{t*} < w^t \cdot x^t$ for any t, then the given data are not consistent with cost-minimizing behavior for any family of production-possibilities sets $\{S^t\}$ satisfying Conditions I above. Note that Conditions I are consistent with there being increasing returns to scale when all inputs and outputs are varied.

My 1979 paper shows how Test 2 can be modified to deal with cases when S^t is a convex set or when S^t is a convex cone and indicates how reasonable measures for the shifts (if any) in the production-possibilities set can be constructed. Additional tests are developed in the paper by Parkan and myself.

V. Comparison of Alternative Approaches

Each of the four methods outlined above for constructing measures of changes in total factor productivity or "exogenous" shifts in the technology involves an approximation error. This is reasonable, since given only market data on prices and quantities, we cannot hope to determine the true production-possibilities sets S^t completely.

The primary advantage of the econometric approach is that it generates estimators for the underlying true production functions f^t (or production-possibilities sets S^t), an advantage that it shares with the nonparametric approach. The main disadvantage of the econometric approach is that it becomes unworkable if there are very large numbers of inputs and outputs.

The primary advantage of the Divisia and exact index number approaches is that they can be implemented even if the number of inputs and outputs is large. The main disadvantage of the Divisia approach is that it does not lead to a definite formula for the shift in the technology, since there are many ways of approximating continuous time derivatives by discrete differences. However, recent work by P. K. Trivedi indicates a promising approach to this problem: use well-known formulae developed in the numerical analysis literature in order to approximate the derivatives optimally in some sense using various finite difference and moving average formulae.

The primary disadvantage of the exact index number approach is that we must assume a very specific functional form for the underlying cost function. On the other hand, the nonparametric approach is free from this disadvantage, but it is computa-

tionally more complex (T linear programming problems have to be solved) and the measures of the shifts in technology that the nonparametric approach generates are only lower bounds to the true shifts.

VI. Special Problems Associated with Capital

The above discussion assumed that the producer sold his vector of outputs produced during period t, y^t, at the prices p^t and rented a vector of inputs x^t at the rental price vector w^t. However, most capital goods cannot readily be *rented*; they must be *purchased*. The fact that a unit of capital purchased during a period lasts longer than that period creates a great number of practical and theoretical problems, some of which we enumerate below.

The simplest way in which the static model of production outlined in Sections I–IV above can be modified to deal with durable inputs is to assume that each durable input or type of capital used by the firm is both an input and an output. More specifically, we could assume that the firm solves the following period t Hicksian profit-maximization problem:

$$(12)\quad \max_{y,x,k,k^*}\left\{ p^t\cdot y + \frac{q^{t+1}\cdot k^*}{1+r} - w^t\cdot x - q^t\cdot k : (y,x,k,k^*)\in S^t\right\}$$

where y is an output vector and p^t is the corresponding price vector, x is a (nondurable) input vector and w^t is the corresponding "wage" vector, k is a vector of durable inputs purchased at the beginning of period t at the prices q^t, k^* is a vector of depreciated durable inputs that will be available to the firm at the beginning of the subsequent period, q^{t+1} is the vector of durable input prices that the firm anticipates will prevail during period t+1, r is an appropriate discount rate, and S^t is the firm's period t feasible set of inputs and outputs. Essentially, we have simply augmented our list of outputs to include depreciated components of the capital stock that will be available next period, and hence the revenue generated by selling the depreciated capital stock must be discounted to reflect the fact that the revenue will not be available until the next period. With the above changes, it would appear that our previous approaches to the measurement of total factor productivity could readily be implemented.

However, appearances can be deceiving. I list below some of the complications which arise when we attempt to implement models based on the Hicksian intertemporal profit-maximization problem (12).

(i) What is relevant in (12) is q^{t+1}, the vector of spot prices for capital stock components that the producer *anticipates* will prevail in the following period. As outside observers, we have no way of actually knowing what these anticipations are.

(ii) What is the relevant *discount rate r*? Actually, this question is closely related to the additional problems which occur when the highly simplified model (12) is expanded to model the effects of the corporation tax, depreciation allowances, and the tax treatment of dividend payments. There is a vast body of literature on these complications which I cannot hope to summarize here.

(iii) Since the anticipated prices q^{t+1} are uncertain, the firm's attitude towards *uncertainty* must be modelled. This leads us into recent developments in the financial theory of the firm, which again is too large an area to summarize here.

(iv) The treatment of *direct and indirect* taxes must be correctly modelled. I have already mentioned that direct taxes such as the corporation profits tax create complications, but the treatment of various indirect taxes such as property taxes, Social Security taxes, and sales taxes can be tricky. Property and Social Security taxes should be added on to the price of the relevant input while a sales tax on the output of the firm should not. See Gollop and Jorgenson on these points.

(v) What should be included in "capital"? In the national accounting systems presently in use, capital consists of produced durable inputs. However, from the viewpoint of the theory of production, it is evident that natural resource stocks and

land should also be included in the list of durable inputs. In addition, stocks of inventories and goods in process should be included. However, the practical problem here is how to obtain data on stocks of resources and inventories when it is not available in our current accounting framework.

(vi) In addition to the problem of obtaining *quantity data* for components of the capital stock, there is the associated problem of obtaining *price data*, particularly for goods in process and for resource stocks, where there are no natural market prices.

VII. Conclusion

This paper presented a brief summary of recent developments in the theory of productivity measurement from the viewpoint of production theory. It should be mentioned that the approaches outlined above can be extended to noncompetitive situations, such as monopolistic behavior (see Robert Allen) or firm behavior under regulatory constraint (see my 1979 paper), usually at the cost of requiring additional information.

I conclude by noting that this discussion of technical change has assumed that it is exogenous. In reality, it is at least partly endogenous, being influenced by past, current, and anticipated future changes in relative prices. As usual, the literature on induced technical change is too large to be summarized here; moreover, its integration into my four approaches to the measurement of total factor productivity is a task for future research.

REFERENCES

S. N. Afriat, (1972a) "Efficiency Estimation of Production Function," *Int. Econ. Rev.*, Oct. 1972, *13*, 568–98.

______, (1972b)"The Theory of International Comparisons of Real Income and Prices," in D. J. Daly, ed., *International Comparisons of Prices and Output*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 37, New York 1972.

R. C. Allen, "Accounting for Price Changes," disc. paper no. 78–14, Univ. British Columbia 1978.

E. R. Berndt and D. O. Wood, "Technology, Prices and the Derived Demand for Energy," *Rev. Econ. Statist.*, Aug. 1975, *57*, 259–68.

H. P. Binswanger, "The Measurement of Technical Change Biases with Many Factors of Production," *Amer. Econ. Rev.*, Dec. 1974, *64*, 964–76.

R. S. Brown, D. W. Caves, and L. R. Christensen, "Modelling the Structure of Cost and Production for Multiproduct Firms," *Southern Econ. J.*, July 1979, *46*, 256–70.

D. W. Caves, L. R. Christensen, and J. A. Swanson, "Productivity in U.S. Railroads 1951–1974," *Bell J. Econ.*, forthcoming.

L. R. Christensen, D. Cummings, and D. W. Jorgenson, "Economic Growth, 1947–1973: An International Comparison," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement*, Chicago 1980.

______, **D. W. Jorgenson, and L. J. Lau,** "Conjugate Duality and the Transcendental Logarithmic Production Function."*Econometrica*, July 1971, *39*, 255–56.

M. Denny, M. Fuss, and L. Waverman, "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries with an Application to Canadian Telecommunications," paper presented at the Conference on Productivity Measurement in Regulated Industries, Univ. Wisconsin-Madison, Apr. 30–May 1, 1979.

W. E. Diewert, "Applications of Duality Theory," in Michael Intriligator and David A. Kendrick, eds., *Frontiers of Quantitative Economics*, Vol. II, Amsterdam 1974.

______, "Exact and Superlative Index Numbers," *J. Econometrics*, May 1976, *4*, 115–45.

______, "Aggregation Problems in the Measurement of Capital," disc. paper no. 77-09, Univ. British Columbia 1977.

______, "Superlative Index Numbers and Consistency in Aggregation," *Econometrica*, July 1978, *46*, 883–900.

______, "The Theory of Total Factor Productivity Measurement in Regulated Industries," paper presented at the Conference on Productivity Measurement in Regulated Industries, Univ. Wisconsin-Madison, Apr. 30–May 1, 1979.

______ and **C. Parkan**, "Linear Programming Test of Regularity Conditions for Production Functions," disc. paper no. 79-01, Univ. British Columbia, Jan. 1979.

M. J. Farrell, "The Measurement of Productive Efficiency," *J. Royal Statist. Soc.*, Part 3, 1957, *120*, 253–82.

Frank Gollop and Dale W. Jorgenson, *U.S. Economic Growth: 1948–1973*, Chicago 1980.

W. M. Gorman, "Measuring the Quantities of Fixed Factors," in J. N. Wolfe, ed., *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, Chicago 1968, pp. 141–72.

G. Hanoch and M. Rothschild, "Testing the Assumptions of Production Theory: A Nonparametric Approach," *J. Polit. Econ.*, Mar./Apr. 1972, *80*, 256–75.

J. R. Hicks, *Value and Capital*, Oxford 1946.

D. W. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Rev. Econ. Stud.*, July 1967, *34*, 249–83.

T. Kloek, "On Quadratic Approximations of Cost of Living and Real Income Index Numbers," rept. no. 6710, Econometric Inst., Netherlands Sch. Econ., Rotterdam 1967.

L. J. Lau, "On Exact Index Numbers," *Rev. Econ. Statist.*, Feb. 1979, *61*, 73–82.

D. McFadden, "Cost, Revenue and Profit Functions," in Melvyn Fuss and his, *Production Economics: A Dual Approach to Theory and Applications*, Vol. 1, Amsterdam 1978, pp. 3–109.

R. A. Pollak, "The Theory of the Cost of Living Index," disc. paper no. 11, Office of Prices and Living Conditions, Bureau of Labor Statistics, Washington 1971.

P. A. Samuelson and S. Swamy, "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *Amer. Econ. Rev.*, Sept. 1974, *64*, 566–93.

J. D. Sargan, "Production Functions," in P. R. G. Layard et al., eds., *Qualified Manpower and Economic Performance*, London 1971.

Ronald W. Shephard, *Theory of Cost and Production Functions*, Princeton 1970.

R. M. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

Henri Theil, *Economics and Information Theory*, Amsterdam 1967.

______, "On the Geometry and the Numerical Approximation of Cost of Living and Real Income Indices," *De Economist*, Sept./Oct. 1968, *116*, 677–89.

P. K. Trivedi, "Some Discrete Approximations to Divisia Integral Indices," unpublished manuscript, Australian Nat. Univ., Apr. 1978.

Welfare Comparison under Exact Aggregation

By DALE W. JORGENSON, LAWRENCE J. LAU, AND THOMAS M. STOKER*

The objective of this paper is to present a new econometric model of aggregate consumer behavior in the United States and to apply this model to the analysis of impacts of alternative economic policies on the welfare of individual consuming units. The model incorporates time-series data on quantities consumed, prices, the level and distribution of income, and demographic characteristics of the population. It also incorporates cross-section data on the allocation of consumer expenditures for households with different demographic characteristics.

Our econometric model is based on the theory of exact aggregation developed by Lau (1977a, c). This theory makes it possible to dispense with the notion of a representative consumer in constructing models of aggregate consumer behavior. One of the most remarkable implications of Lau's theory of exact aggregation is that systems of demand functions for individuals with common demographic characteristics can be recovered uniquely from the system of aggregate demand functions. Using the individual demand functions we can analyze the impact of economic policy on consumer welfare.

I. Econometric Model

We assume that each consuming unit has an indirect utility function that is homogeneous of degree zero in prices and income, nonincreasing in prices and nondecreasing in income, and quasi convex in prices. To incorporate differences in individual preferences, we allow the indirect utility function to depend on attributes such as demographic characteristics that vary among individuals.

We assume that there are K consuming units, indexed by $k=1,2\ldots K$, and N goods, indexed by $n=1,2\ldots N$. The quantity of the nth good has a price p_n, assumed to be the same for all consuming units at a given point of time. In our econometric model there are five commodity groups: 1) Energy: Electricity, gas, heating oil, gasoline, and so on; 2) Agricultural products: Food, clothing, and related expenditures; 3) Consumer services: Personal services, entertainment, insurance and so on; 4) Capital services: Service flow from consumer durables and housing; 5) Trade and transportation: Purchases of transportation services and trade and transportation margins on all other goods and services. Total expenditure by the kth unit on all five commodity groups is $M_k=\Sigma p_n x_{nk}$. The expenditure share of the nth good in the budget of the kth consuming unit is $w_{nk}=p_n x_{nk}/M_k$.

In our econometric model, the consuming units are households. To allow for differences in preferences among consuming units, we allow the indirect utility function for the kth unit to depend on a vector of attributes, say A_k; each attribute is represented by a dummy variable equal to unity when the consuming unit has the corresponding characteristic and zero otherwise. In our econometric model there are five groups of attributes: 1) Family size: 1, 2, 3, 4, 5, 6, and 7 or more persons; 2) Age of head: 15–24, 25–34, 35–44, 45–54, 55–64, 65 years and over; 3) Region of residence: Northeast, North Central, South, and West; 4) Race: white, nonwhite; 5) Type of residence: urban, rural. Each consuming unit is assigned one of the attributes in each of the five groups.

To represent our econometric model we require the following additional notation:

$w_k=(w_{1k}, w_{2k}\ldots w_{Nk})$: vector of expenditure shares for the kth consuming unit

$ln\frac{p}{M_k}=(ln\frac{p_1}{M_k}, ln\frac{p_2}{M_k}\ldots ln\frac{p_N}{M_k})$: vector of

*Professor of economics, Harvard University, professor of economics, Stanford University, and assistant professor of economics, Sloan School of Management, Massachusetts Institute of Technology, respectively.

logarithms of ratios of prices to total expenditures for the kth consuming unit

$lnp=(lnp_1, lnp_2 \ldots lnp_N)$: vector of logarithms of prices

We assume that the kth consuming unit allocates its expenditures in accord with the transcendental logarithmic or translog[1] indirect utility function, say V_k:

$$ln V_k = \left(ln \frac{p}{M_k}\right)' \alpha_p + \frac{1}{2}\left(ln \frac{p}{M_k}\right)' \times \beta_{pp}\left(ln \frac{p}{M_k}\right) + \left(ln \frac{p}{M_k}\right)' \beta_{pA} A_k$$

where α_p, β_{pp}, and β_{pA} are matrices of parameters that are constant over all consuming units.

The expenditure shares of the kth consuming unit can be derived by Roy's Identity:[2]

$$w_k = \frac{1}{D_k}\left(\alpha_p + \beta_{pp}\, ln \frac{p}{M_k} + \beta_{pA} A_k\right)$$

where the denominator D_k takes the form

$$D_k = i'\alpha_p + i'\beta_{pp}\, ln \frac{p}{M_k} + i'\beta_{pA} A_k$$

where i is a vector of ones.

The conditions for exact aggregation are that the individual demand functions are linear in functions of the attributes A_k and total expenditure M_k.[3] These conditions will be satisfied if and only if terms involving the attributes and total expenditure do not appear in the denominators, so that $i'\beta_{pp} i = 0$, and $i'\beta_{pA} = 0$. These restrictions imply that the denominators D_k reduce to $D = i'\alpha_p + i'\beta_{pp}\, lnp$, where the subscript k is no longer required, since the denominators are the same for all consuming units.

Aggregate expenditure shares, say w, are obtained by multiplying individual expenditure shares by total expenditure for each consuming unit, adding over all consuming units, and dividing by total expenditure for all units: $w = \Sigma M_k w_k / \Sigma M_k$. The aggregate expenditure shares take the form:

$$w = \frac{1}{D}\left(\alpha_p + \beta_{pp}\, lnp - \beta_{pp} i \frac{\Sigma M_k\, ln M_k}{\Sigma M_k} + \beta_{pA} \frac{\Sigma M_k A_k}{\Sigma M_k}\right)$$

Aggregate expenditure shares depend on prices p. They also depend on the distribution of total expenditure over consuming units through the statistic $\Sigma M_k\, ln M_k / \Sigma M_k$. Finally, they depend on the distribution of total expenditures over demographic groups through the statistics $\Sigma M_k A_k / \Sigma M_k$.

The parameters $\beta_{pp} i$ and β_{pA} can be estimated from cross-section data on expenditures on all commodity groups, total expenditure, and demographic characteristics of all consuming units. Time-series data on prices are required to estimate the remaining parameters of the model. To implement our model of aggregate consumer behavior we pool time-series data for the period 1958–74 with cross-section data for 1972 from the 1972–73 *Survey of Consumer Expenditures* (see M. D. Carlson). We estimate $\Sigma M_k\, ln M_k / \Sigma M_k$ and $\Sigma M_k A_k / \Sigma M_k$ on a time-series basis from the *Current Population Reports*.[4]

II. Policy Analysis

To evaluate the impact of alternative economic policies we employ the compensating variation in total expenditure required for each consuming unit to achieve the same level of utility before and after the policy change. If the compensating variation is

[1]The translog indirect utility function was introduced by Laurits Christensen, Jorgenson, and Lau, and extended to incorporate changes in preferences over time by Jorgenson and Lau.

[2]See René Roy; a detailed review of econometric models of consumer behavior based on Roy's Identity is given by Lau (1977a).

[3]For details see Lau (1977b).

[4]Our time-series are based on data prepared by Jack Faucett Associates for the Federal Preparedness Agency. Numerical estimates of all parameters are given in an appendix to this paper; copies are available from the authors upon request. For details on the estimation procedure, see Thomas Stoker.

TABLE 1—PRICE, INCOME, AND EXPENDITURE PROJECTIONS

	Prices (1972=1.000)									
	Base Case					Case with Decontrol				
	1	2	3	4	5	1	2	3	4	5
1979	2.1170	1.6193	1.6553	1.6000	1.6989	2.1276	1.6193	1.6553	1.6000	1.6989
1980	2.3048	1.7246	1.7878	1.7246	1.8404	2.3537	1.7229	1.7878	1.7246	1.8423
1981	2.4818	1.8313	1.9309	1.8552	1.9957	2.5883	1.8368	1.9348	1.8571	1.9997
1982	2.6698	1.9367	2.0709	1.9858	2.1511	2.8039	1.9484	2.0751	1.9897	2.1598
1983	2.8691	2.0442	2.2034	2.1170	2.3117	3.0072	2.0565	2.2144	2.1276	2.3233
1984	3.0771	2.1555	2.3420	2.2547	2.4818	3.2123	2.1684	2.3561	2.2660	2.4943
1985	3.2772	2.2728	2.4818	2.3941	2.6538	3.4281	2.2864	2.4968	2.4061	2.6671

	Income and Expenditures							
	Base Case				Case with Decontrol			
	I (1979=	M \$17,000)	I (1979=	M \$8,000)	I (1979=	M \$17,000)	I (1979=	M \$8,000)
1979	17000.00	15740.98	8000.00	7407.52	17000.00	15740.98	8000.00	7407.52
1980	18720.93	17334.48	8809.85	8157.39	18758.20	17368.97	8827.39	8173.63
1981	20651.22	19121.79	9718.22	8998.49	20766.33	19228.38	9772.39	9048.65
1982	22354.96	20699.35	10519.98	9740.87	22527.57	20859.18	10601.21	9816.08
1983	24520.38	22704.40	11539.00	10684.42	24705.32	22875.64	11626.03	10765.01
1984	26970.78	24973.32	12692.13	11752.15	27180.49	25167.50	12790.82	11843.53
1985	29450.61	27269.50	13859.11	12832.70	29710.52	27510.16	13981.42	12945.96

Source: Edward Novicky, Scientific Time Sharing Corporation, based on the DRI Quarterly Model of the United States.

negative, the welfare of the consuming unit is increased by the policy change; if the compensating variation is positive, the welfare of the consuming unit is decreased. Differences in compensating variations among consuming units reflect the fact that preference and economic circumstances differ among units.

Under the exact aggregation condition the indirect utility function for each consuming unit takes the form:

$$\ln V_k = \ln p'\left(\alpha_p + \tfrac{1}{2}\beta_{pp}\ln p + \beta_{pA}\right) - \ln M_k D(p)$$

Given the indirect utility function for each unit, we can solve explicitly for the expenditure function:

$$\ln M_k = \frac{1}{D}\left[\ln p'\left(\alpha_p + \frac{1}{2}\beta_{pp}\ln p + \beta_{pA}A_k\right) - \ln V_k\right]$$

The expenditure function gives the minimum expenditure required for the consuming unit to achieve the utility level V_k, given prices p.

To analyze the impact of a change in economic policy on the kth household, we first evaluate the indirect utility function before the change in policy takes place. Suppose the prices are p^0 and expenditure for the kth household is M_k^0. Now, suppose that a change in policy results in prices p^1 and expenditure for the kth household M_k^1. We define the compensating variation in total expenditure for the kth household, say C_k, as the additional expenditure required to achieve the same level of utility as before the change in policy, say V_k^0:

$$C_k = M_k(p^1, V_k^0, A_k) - M_k^1$$

The compensating variation depends on the attributes A_k of the kth consuming unit, on initial expenditure and prices, which enter through the indirect utility function of the kth consuming unit $V_k^0(p^0, A_k, M_k^0)$, and on the prices p^1 and expenditure M_k^1 resulting from the policy change.

TABLE 2—CHANGE IN CONSUMER WELFARE (1979 Income = $17,000)

	Urban White	Urban Nonwhite	Rural White	Rural Nonwhite
Northeast				
1979	−4.14	−3.80	−5.45	−5.11
1980	15.67	17.22	10.34	11.89
1981	33.73	36.60	22.19	25.05
1982	45.00	48.14	31.93	35.08
1983	9.53	13.55	−6.16	−2.14
1984	15.87	19.77	.34	4.25
1985	44.05	48.55	26.18	30.68
South				
1979	−2.79	−2.45	−4.10	−3.75
1980	21.01	22.56	15.68	17.23
1981	47.14	50.00	35.60	38.47
1982	61.05	64.19	47.99	51.14
1983	26.30	30.32	10.63	14.64
1984	32.91	36.81	17.40	21.30
1985	63.61	68.11	45.75	50.25
North Central				
1979	−4.83	−4.48	−6.13	−5.79
1980	12.86	14.41	7.52	9.07
1981	26.95	29.82	15.40	18.27
1982	36.42	39.57	23.35	26.50
1983	1.05	5.07	−14.64	−10.62
1984	7.51	11.41	−8.02	−4.11
1985	34.44	38.95	16.57	21.07
West				
1979	−3.41	−3.06	−4.71	−4.37
1980	18.77	20.32	13.44	14.99
1981	40.91	43.77	29.36	32.23
1982	53.38	56.52	40.32	43.47
1983	18.51	22.53	2.83	6.85
1984	25.03	28.93	9.51	13.41
1985	54.57	59.07	36.71	41.21

TABLE 3—CHANGE IN CONSUMER WELFARE (1979 Income = $8,000)

	Urban White	Urban Nonwhite	Rural White	Rural Nonwhite
Northeast				
1979	−2.47	−2.31	−3.09	−2.92
1980	5.12	5.85	2.61	3.34
1981	9.27	10.62	3.84	5.18
1982	12.14	13.63	5.99	7.47
1983	−2.42	−.53	−9.81	−7.92
1984	.26	2.09	−7.05	−5.21
1985	12.49	14.61	4.07	6.20
South				
1979	−1.84	−1.67	−2.45	−2.29
1980	7.63	8.36	5.12	5.85
1981	15.59	16.94	10.15	11.50
1982	19.70	21.18	13.55	15.04
1983	5.48	7.37	−1.90	−.01
1984	8.28	10.12	.98	2.82
1985	21.70	23.82	13.29	15.41
North Central				
1979	−2.79	−2.63	−3.41	−3.24
1980	3.79	4.52	1.28	2.01
1981	6.08	7.43	.64	1.99
1982	8.10	9.59	1.95	3.43
1983	−6.41	−4.52	−13.80	−11.91
1984	−3.68	−1.84	−10.99	−9.15
1985	7.97	10.09	−.45	1.67
West				
1979	−2.13	−1.96	−2.74	−2.58
1980	6.58	7.31	4.07	4.80
1981	12.65	14.00	7.22	8.56
1982	16.09	17.57	9.94	11.42
1983	1.81	3.70	−5.58	−3.68
1984	4.57	6.41	−2.74	−.90
1985	17.44	19.56	9.03	11.15

We illustrate the application of our model of aggregate consumer behavior by analyzing the impact of a policy of decontrolling the prices of oil products in 1979. We measure the compensating variations in expenditures over a seven-year period for individual consuming units with different demographic characteristics and different income levels. For this purpose we employ projections of prices for the five commodity groups included in our model with and without oil price decontrol for the period 1979–85. We also employ projections of income, labelled I in the tables, and total expenditure M for households with 1979 incomes of $17,000 and $8,000. The projections are given in Table 1 for the base case and the case with decontrol.

To evaluate the impact of oil price decontrol on consumer welfare, we evaluate the compensating variation in expenditure for each consuming unit at prices p^1 and total expenditure M_k^1 with decontrol, and utility level V_k^0 without decontrol. In Table 2 we give the negative of the compensating variation for households with $17,000 of income in 1979, having four members and age of head of household in the range from 35–44 years. Similarly, in Table 3 we give the negative of the compensating variation for households with $8,000 in income in 1979 with the same demographic characteristics. Results are given for consuming units in each of four regions of the United States, for urban vs. rural residents, and for white vs. nonwhite racial groups.

A comparison of compensating variations among groups reveals that benefits of oil price decontrol are proportionately greater for consumer groups with higher incomes. Second, at a given level of income, urban groups generally benefit more than rural groups. Again at a given level of income, nonwhites benefit somewhat more than whites. Finally, households in the South and West regions of the United States benefit more than households in the Northeast and North Central regions. Households in the North Central region benefit the least; rural households in this region with 1979 income of $8,000 appear to be the only group we have examined that fails to benefit from oil price decontrol.

III. Conclusion

The novel feature of our model of aggregate consumer expenditures is that we are able to dispense with the notion of a representative consumer by using the theory of exact aggregation. This theory makes it possible to incorporate prices, the level and distribution of income, and the demographic characteristics of the population into projections of aggregate expenditure patterns. The theory also makes it possible to recover systems of individual demand functions and to analyze the impact of economic policy on individual consumer welfare. In this paper we have not attempted to combine changes in welfare for different households into an overall measure of the change in social welfare. We plan to develop this extension of our methodology in future research.

REFERENCES

M. D. Carlson, "The 1972–1973 Consumer Expenditure Survey," *Mon. Labor Rev.*, Dec. 1974, *97*, 16–23.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Transcendental Logarithmic Utility Functions," *Amer. Econ. Rev.*, June 1975, *65*, 367–83.

D. W. Jorgenson and L. J. Lau, "The Structure of Consumer Preferences," *Annals Econ. Soc. Measure.*, Jan. 1975, *4*, 49–101.

L. J. Lau, (1977a) "Complete Systems of Demand Functions Through Duality," in Michael D. Intriligator and David A. Kendrick, eds., *Frontiers of Quantitative Economics*, Vol. III, Amsterdam 1977, pp. 59–85.

______, (1977b) "Existence Conditions for Aggregate Demand Functions: The Case of Multiple Indexes," tech. rept. no. 249, IMSSS, Stanford Univ., Oct. 1977.

______, (1977c) "Existence Conditions for Aggregate Demand Functions: The Case of a Single Index," tech. rept. no. 248, IMSSS, Stanford Univ., Oct. 1977.

René Roy, *De l'Utilité: Contribution à la Théorie des Choix*, Paris 1943.

T. M. Stoker, "Aggregation Over Individuals and Demand Analysis," unpublished doctoral dissertation, Harvard Univ. 1979.

Jack Faucett Associates, *Development of 35 Order Input-Output Tables, 1958–1974, Final Report*, Washington, Oct. 1977.

U.S. Bureau of the Census, *Current Population Reports, Series P-60, Consumer Income*, various issues.

Group Cost-of-Living Indexes

By ROBERT A. POLLAK*

When households have different consumption patterns, whose cost of living should an actual price index represent? This issue was first raised by J. L. Nicholson and S. J. Prais in the 1950's. Both made essentially the same point: official price indexes give each household's consumption pattern "an implicit weight proportional to its total expenditures" (see Nicholson, p. 540). Prais calls such an index "plutocratic," and both Nicholson and Prais suggest an alternative "democratic price index" which gives all households equal weight.

A "group cost-of-living index" is an index that measures the impact of price changes on the welfare of a group or population of households. To define such an index requires an explicit or implicit concept of "the welfare of a group," and hence requires interpersonal comparison and distributional judgments. Since group indexes such as the Consumer Price Index play an important role in our perception of inflation and the formation of macro-economic policy and are used to escalate wages and Social Security benefits, they have significant effects on government decisions and economic welfare. Despite their intellectual interest and practical importance, however, until recently they have been virtually ignored by index number theorists. The theory of the cost-of-living index (*CLI*) provides a generally accepted framework for measuring the impact of price changes on the welfare of a particular household. This paper extends the *CLI* concept to groups and discusses which questions require group indexes and which do not. I begin by introducing some notation and terminology in the context of household *CLI*s.

A household's *CLI* is the ratio of the expenditures required to attain a particular base indifference curve in two price situations. Suppose there are n goods and S households, and denote the preference ordering of the rth household by R^r. The base indifference curve can be identified by a goods collection, X^{ro}, which lies on it. The "expenditure function," $E^r(P,X^{ro},R^r)$, shows the minimum expenditure required to attain the base indifference curve at prices P. The *CLI* of the rth household, $I^r(P^a,P^b,X^{ro},R^r)$, is the ratio of the minimum expenditure required to attain the base indifference curve at prices P^a (comparison prices) to that required at prices P^b (reference prices).

Except in very special cases, the value of the *CLI* depends on the base indifference curve at which it is evaluated; as successively higher base indifference curves are specified, one would expect the prices of "luxuries" to become more important relative to the prices of "necessities."[1] Hence, it is convenient to regard the *CLI* as a function of the base indifference curve rather than as a single number corresponding to a particular base. Thus, instead of offering guidance in choosing an appropriate base indifference curve, theory suggests that there is no need to choose.

To construct the exact *CLI*, an investigator needs to know the household's preferences. Lacking this knowledge, he must fall back on indexes which require less information and which are upper bounds on the exact index. The "Laspeyres index," $J^r(P^a,P^b,X^{rb})$, is the ratio of the cost of purchasing the reference period consumption basket at comparison prices to its cost

*University of Pennsylvania. This research was supported in part by the National Science Foundation, the U.S. Bureau of Labor Statistics, and the National Institutes of Health, none of which are responsible for the views expressed.

[1]The value of the cost-of-living index is independent of the base indifference curve at which the index is evaluated if and only if the indifference map is homothetic to the origin, or, equivalently, if and only if all income elasticities are unity.

at reference prices:

$$J^r(P^a, P^b, X^{rb}) = \frac{\sum_{k=1}^{n} p_k^a x_k^{rb}}{\sum_{k=1}^{n} p_k^b x_k^{rb}}$$

Provided the household's reference period consumption pattern was optimal (according to its own preferences), the Laspeyres index is an upper bound on its cost-of-living index evaluated at the reference period indifference curve: $I^r(P^a, P^b, X^{rb}, R^r) \leqslant J^r(P^a, P^b, X^{rb})$.

To measure the impact of price changes on the welfare of a group of households, I consider two basic group *CLI*s, the "social *CLI*" and the "democratic *CLI*," and several related indexes. Since differences in family size and other demographic characteristics may require special treatment, it is convenient to begin with a group of households with identical demographic profiles. Distributional judgments play a crucial role in constructing an index for such a group, a role most easily seen in the context of Laspeyres-type indexes. Such indexes are ratios of the costs of particular collections of goods at comparison prices to their costs at reference prices. When households have different reference period consumption patterns—reflecting underlying differences in tastes or incomes—some implicit or explicit criterion is required to combine them into a single "social consumption pattern" on which to base the index. Nicholson and Prais, who consider only Laspeyres-type indexes, view the issue as one of weighting the consumption patterns of different households, but distributional judgments also play a crucial role in "exact" group indexes.

The social *CLI* is a group index based on a Bergson-Samuelson social welfare function, in the same way that a household's *CLI* is based on its own utility function; the social welfare function provides an explicit criterion for balancing the gains of some households against losses of others. The democratic *CLI* is not based on any explicit criterion for trading off the welfare of one household against that of another; it is defined as the unweighted average of the *CLI*s of the covered households (i.e., the households in the group). To construct either of these indexes requires knowledge of household preferences; in addition, the social *CLI* requires the investigator to specify his Bergson-Samuelson social welfare function.

Since an investigator is unlikely to have enough information to construct either of these exact indexes, he needs indexes which bound them and can be calculated with less information. Bounds on the democratic *CLI* are relatively easy to construct, but because the social *CLI* is intimately related to a particular Bergson-Samuelson social welfare function, useful bounds on it can be established only in very special cases.

I. The Social *CLI*[2]

The social *CLI* is a group index based on a Bergson-Samuelson social welfare function. To construct it, an investigator must know each household's preferences and must express his distributional judgments in the form of a Pareto-inclusive social welfare function. I denote the welfare function by $W(\chi) = \Lambda[U^1(X^1), \ldots, U^S(X^S)]$ (and the corresponding social preference ordering by R) where $\chi = (X^1, \ldots, X^S)$ is the $n \times s$ dimensional "social consumption vector" showing every household's consumption of every good.

The "social expenditure function," $E(P, \chi^0, R)$, shows the minimum total expenditure required to attain the indifference curve of χ^0 at prices P. The expenditure function represents a "thought experiment" in which expenditure is distributed among households so as to minimize the cost of attaining the specified indifference curve of the social welfare function.[3] The social *CLI*, $I(P^a$,

[2]The material summarized in this section is drawn from my 1976 paper.

[3]A "constrained social *CLI*" can be defined if the distribution of expenditure among households must satisfy additional restrictions beyond the requirement that $\Sigma \mu_s = \mu$.

$P^b, \chi^0, R)$, is the ratio of the minimum expenditure required to attain an indifference curve of the social welfare function at prices P^a to that required at prices P^b:

$$I(P^a,P^b,\chi^0,R)=\frac{E(P^a,\chi^0,R)}{E(P^b,\chi^0,R)}$$

The information needed to construct the exact social *CLI* is unlikely to be available. There are, however, two special cases in which the exact social *CLI* can be constructed with much less information. In a "maximizing society"—one which distributes expenditure among households so as to maximize a social welfare function—the market demand functions contain enough information to construct the exact social *CLI* corresponding to society's welfare function. The usefulness of this result depends on whether society distributes expenditure among households so as to maximize a social welfare function, and, if it does, on whether the investigator adopts that welfare function as the basis for his social *CLI*. In an "independent society" —one in which household preferences are such that the market demand functions are independent of the distribution of expenditure among households—it is well known that the market demand functions can be derived from the preference ordering of a "representative household" and that the corresponding indifference map can be constructed from the market demand functions. In my discussion paper, I show that in an independent society the social *CLI* corresponding to any Pareto-inclusive social welfare function is the index corresponding to the preference ordering of the representative household. Hence, disagreement among investigators with different social welfare functions is limited to the selection of an appropriate base indifference curve from the commonly accepted indifference map of the representative household. Because both the maximizing society and the independent society are very special cases, neither provides a general procedure for constructing the exact social *CLI*.

Lacking the information required to calculate the exact social *CLI*, an investigator may be able to calculate indexes which are upper bounds on it. However, the assumptions required to place bounds on the social *CLI* are much less likely to be satisfied than those required to place analogous bounds on a household's *CLI*. I consider two bounding indexes, the "social Laspeyres index" and the "Scitovsky-Laspeyres index."

The social Laspeyres index, $J(P^a,P^b,X^b)$, is defined as the ratio of the expenditure required to purchase the reference period social consumption basket, $X^b=\Sigma X^{sb}$, at comparison prices to its cost at reference prices:

$$J(P^a,P^b,X^b)=\frac{\sum_{k=1}^{n} p_k^a x_k^b}{\sum_{k=1}^{n} p_k^b x_k^b}=\frac{\sum_{s=1}^{S}\sum_{k=1}^{n} p_k^a x_k^{sb}}{\sum_{s=1}^{S}\sum_{k=1}^{n} p_k^b x_k^{sb}}$$

It can be rewritten as a weighted average of household Laspeyres indexes, where the weights are the household's reference period expenditure shares:

$$J(P^a,P^b,X^b)$$
$$=\sum_{s=1}^{S}\omega^s(P^b,\chi^b)J^s(P^a,P^b,X^{sb})$$

$$\omega^r(P^b,\chi^b)=\frac{\sum_{k=1}^{n} p_k^b x_k^{rb}}{\sum_{s=1}^{S}\sum_{k=1}^{n} p_k^b x_k^{sb}}=\frac{\mu_r^b}{\mu^b}$$

The Scitovsky-Laspeyres index, $J^*(P^a,P^b, \chi^b, R^1,\dots,R^S)$, is defined as the ratio of the total expenditure required to enable each household to attain its reference period indifference curve at comparison prices to that required at reference prices; it can be rewritten as a weighted average of household *CLI*s where the weights are again

reference period expenditure shares:

$$J^*(P^a,P^b,\chi^b,R^1,\ldots,R^S)$$
$$=\sum_{s=1}^{S}\omega^s(P^b,\chi^b)I^s(P^a,P^b,X^{sb},R^s)$$

The social Laspeyres index is an upper bound on the social *CLI* for an appropriately chosen base indifference curve in the independent society and in the maximizing society, when the investigator adopts society's social welfare function. The Scitovsky-Laspeyres index is an upper bound on the social *CLI* in a maximizing society when the investigator adopts society's welfare function. However, it is not generally possible to place bounds on the exact social *CLI* without some assumption which guarantees that the reference period consumption pattern is optimal in terms of the social welfare function on which the index is based. Without such an assumption, we cannot construct indexes which provide bounds on the exact social *CLI* analogous to the Laspeyres bound on the household's *CLI*.

II. The Democratic *CLI* and the Axiomatic Approach

Nicholson and Prais both define the democratic price index, $J^D(P^a,P^b,\chi^b)$, to be a weighted average of "price relatives," where the weight of each item is the mean of its shares in the consumption patterns of the households in the group:

$$J^D(P^a,P^b,\chi^b)=\sum_{k=1}^{n}w_k^D(P^b,\chi^b)\left(\frac{p_k^a}{p_k^b}\right)$$

where

$$w_i^D=\frac{1}{S}\sum_{s=1}^{S}w_i^s(P^b,\chi^b)$$

and $$w_i^s(P^b,\chi^b)=\frac{p_i^b x_i^{sb}}{\sum_{k=1}^{n}p_k^b x_k^{sb}}$$

The index can be rewritten as an unweighted average of household Laspeyres indexes:

$$J^D(P^a,P^b,\chi^b)=\frac{1}{S}\sum_{s=1}^{S}J^s(P^a,P^b,X^{sb})$$

In contrast, the social Laspeyres index implicitly weights each household's Laspeyres index by its total expenditure. Thus, in the democratic price index, every household counts equally, while in the "plutocratic" social Laspeyres index, every dollar of expenditure counts equally.

The democratic price index, like the social Laspeyres index, is a fixed-weight index and hence fails to take account of the willingness of households to substitute one good for another in response to relative price changes. I define the democratic *CLI*, $I^D(P^a,P^b,\chi^0)$, to be an unweighted average of household *CLI*s:

$$I^D(P^a,P^b,\chi^0)=\frac{1}{S}\sum_{s=1}^{S}I^s(P^a,P^b,X^{s0},R^s)$$

Unlike the democratic price index, the democratic *CLI* permits some substitution in response to relative price changes; in particular, it permits "within household" substitution, but unlike the social *CLI*, it does not permit "between household" substitution. The democratic *CLI* is analogous to the Scitovsky-Laspeyres index: both are fixed-weight indexes which are averages of household *CLI*s. Just as the social Laspeyres index is an upper bound on the Scitovsky-Laspeyres index, the democratic price index is an upper bound on the democratic *CLI*: $I^D(P^a,P^b,\chi^b)\leq J^D(P^a,P^b,\chi^b)$.[4]

[4]John Muellbauer, in an interesting unpublished paper, proposes the class of "homogeneous social price indexes," $J^M(P^a,P^b,\chi^b,\beta)$,

$$J^M(P^a,P^b,\chi^b,\beta)=\sum_{s=1}^{S}(\mu_s^b)^{1+\beta}J^s(P^a,p^b,X^{sb})\bigg/\sum_{t=1}^{S}(\mu_t^b)^{1+\beta}$$

This class depends on a parameter β, which Muellbauer interprets as a measure of aversion to in-

The democratic *CLI* and the social *CLI* represent alternative approaches to defining group *CLI*s. Although the democratic *CLI* coincides with the social *CLI* or the Scitovsky-Laspeyres index in certain exceptional cases, it is neither a generalization nor a specialization of these indexes. The democratic *CLI* is not an approximation to or a bound on some more fundamental index, but is itself a basic index. Although it cannot be interpreted as a *CLI* corresponding to an underlying social welfare function, the meaning of the democratic *CLI* is unambiguous: it is an unweighted average of the *CLI*s of the households in the group, each evaluated at its reference period indifference curve. That is, different households require different percentage increases in their reference period expenditure levels to enable them to maintain their reference period indifference curves under the comparison price regime; the democratic *CLI* is defined to be the mean of these required percentage adjustments.

The axiomatic approach to price index construction—an outgrowth of the Irving Fisher test tradition—has been largely eclipsed by the modern, preference-based theory of the *CLI*. As a framework for constructing indexes for particular households, the preference approach is unlikely to be displaced by a revival of the test approach. However, as a framework for constructing group indexes, the axiomatic approach deserves more attention than it has thus far received, since the advantages of the preference approach are greatly attenuated when we turn from household to group indexes.

equality; when $\beta=-1$, all households receive equal weight (equal "votes" in Muellbauer's terminology) and the index coincides with the democratic price index; when $\beta=0$, it coincides with the social Laspeyres index. If Muellbauer's parametric weighting procedure is applied to household *CLI*s instead of household Laspeyres indexes, it yields a class of "homogeneous social *CLI*s" which permit within household substitution in response to relative price changes. This class includes the democratic *CLI* and the Scitovsky-Laspeyres index as special cases. For each β, the price indexes are upper bounds on the corresponding *CLI*s.

III. Comparisons without Group Indexes

Most questions about the distributional effects of price changes can be answered without group indexes. This section discusses which questions require group indexes and which do not.

We can always compare the impact of price changes on the welfare of different households by comparing their (separate) *CLI*s. Such comparisons do not require balancing one household's interests against another's, and thus can be made without distributional judgments. Group indexes and distributional judgments are necessary only if we require a single summary measure which reflects the impact of price changes on the entire group.

Suppose, for example, we want to compare the impact of price changes on the welfare of households with two children and those with three children.[5] If we know their preferences, we can construct their separate *CLI*s; however, it is not clear how to compare them, since, except in special cases, the indexes are not single numbers but functions whose values depend on the base indifference curves at which they are evaluated. If we insist on comparing *CLI functions*, there is little to say: theory offers no guidance for such comparisons. If we reinterpret the problem as one of comparing the *values* of the functions at a particular pair of base indifference curves, then we avoid this difficulty; however, theory offers no guidance for selecting an appropriate pair of base indifference curves. If we specify the base indifference curve for two-child households as that corresponding to an expenditure level of \$12,000 at reference prices, and the base for three-child households as that corresponding to \$13,000, then a straightforward comparison can be made; however, instead of comparing two-child households and three-child households, we are now comparing \$12,000 two-child

[5]This example involves two types of households and taste differences related to demographic characteristics, but it generalizes readily to many household types and any kind of taste differences.

households and \$13,000 three-child households.[6]

Discussions of "whose *CLI*" have emphasized differences in consumption patterns resulting from income differences and their effect on Laspeyres-type indexes. We can compare the impacts of price changes on rich and poor households with identical tastes by comparing the value of the *CLI* corresponding to different base indifference curves; the formulation of the question in terms of "rich" and "poor" households implies a particular pair of base indifference curves, or at least a range of appropriate pairs. To formalize this notion, I define the "common *CLI*" for a group of households with identical tastes, $I^c(P^a,P^b,X^0,R)$, to be the *CLI* corresponding to their common preference ordering, R. The common *CLI* coincides with the social *CLI* if the investigator's social welfare function exhibits "equal concern" for all households—that is, if it treats all households symmetrically. It coincides with the democratic *CLI* if all households have identical reference period consumption patterns.

When household preference orderings are not known, exact indexes cannot be constructed. However, we can construct Laspeyres indexes from data on the reference period consumption patterns. Robert Michael reports household-specific Laspeyres indexes for more than 11,000 households and discusses the sources of variations in these indexes.

To summarize: Group *CLI*s measure the impact of price changes on a group of households, but different group indexes answer different questions, and the selection of a particular index must be determined by its purpose. Since the information required to construct exact indexes is unlikely to be available, we are forced to consider bounds; however, even the construction of bounds presents difficulties for indexes such as the social cost-of-living index, which is based on a Bergson-Samuelson social welfare function. Most questions concerning the distributional impact of price changes can be answered by comparing household *CLI*s, and do not call for group indexes at all. Group indexes are needed only if we require a summary measure of the welfare impact of price changes on the entire group. In short: distributional comparisons do not require group indexes, but group indexes require distributional judgments.

[6]This procedure would be somewhat more attractive if we could determine the level of expenditure which would make a household with three children as well off as one with two children and \$12,000, since this would enable us to compare the welfare impact of price changes on households at the same "standard of living." I do not believe that it is possible to use household budget data to establish the required correspondence between the indifference curves on one map and those on another, although professional opinion is divided on this question. See Muellbauer for a defense of the use of "equivalence scales" to make such comparisons, and the paper by Wales and myself for an attack on it.

REFERENCES

R. T. Michael, "Variations Across Households in the Rate of Inflation," *J. Money, Credit, Banking,* Feb. 1979, *11*, 32–46.

J. Muellbauer, "The Political Economy of Price Indices," disc. paper no. 22, Birkbeck College, Mar. 1974.

J. L. Nicholson, "Whose Cost off Living?," *J. Royal Statist. Soc., Part* 4, 1975, *138*, 540–42.

S. J. Prais, "Whose Cost of Living?," *Rev. Econ. Stud.*, Feb. 1959, *26*, 126–34.

R. A. Pollak, "The Social Cost of Living Index," disc. paper no. 378, Univ. Pennsylvania, June 1976.

______, **and T. J. Wales,** "Welfare Comparisons and Equivalence Scales," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 216–21.

DISCUSSION

LAURITS R. CHRISTENSEN, University of Wisconsin-Madison: W. E. Diewert has provided an excellent overview of the theoretical underpinnings of several approaches to the estimation of total factor productivity. I will discuss the approaches from the point of view of one who wishes to use the best available techniques for estimating productivity growth.

Diewert emphasizes the differences among the approaches and treats them as substitutes. The practitioner would be well-advised to recognize the similarities of the approaches and to view them as complementary. Productivity growth, as defined by Diewert, is a shift in the production function over time. All of the approaches discussed by Diewert presume the existence of a well-behaved production function and attempt to disentangle shifts in the production function from movements along the function. Since the objective is the same in each case, it is desirable to bring multiple methods to bear on the problem, rather than relying on the findings of a single approach. A good productivity study might well read like a good detective story: clues from a variety of approaches can supplement each other and provide important cross checks in order to shed light on the true (but not directly observable) state of affairs. Results from alternative approaches are particularly important when evaluating the econometric approach which is complicated and contains the most potential pitfalls. Armed with clues from the other approaches, the investigator will be in a much better position to interpret results from the econometric approach.

Diewert distinguishes four approaches, but in view of the very close relation between the Divisia and exact index number approaches, it is useful to combine them under the heading of "the index number approach." This approach has received considerable attention in recent years due, in no small part, to Diewert's 1976 article which introduced the concept of superlative index number procedures. This work made it possible to choose among index number procedures on the basis of the flexibility which they permit in the underlying structure of production. Thus, the traditional Paasche and Laspeyres indexes could be abandoned in favor of procedures which do not unduly constrain the production structure. The nonparametric approach has received less attention than the index number approach, partly because it is still in the experimental stage, but also because it sacrifices much of the simplicity associated with the index number approach.

Is there a logical order in which to apply multiple methods? It is my view that best practice in productivity estimation is to begin by employing an unrestrictive index number procedure and to conclude by using the econometric approach, employing a flexible functional form. It is sufficient to rely solely on the index number result only if the assumptions of constant returns to scale and full competitive equilibrium are deemed satisfactory. Otherwise, estimated productivity growth will reflect such factors as scale diseconomies and changes in capacity utilization—which are movements along the production function.

While these problems may suggest that the index number approach be abandoned entirely, such a conclusion is unwarranted. Most of the effort required to obtain index number estimates is also required to obtain econometric estimates. Consideration of index number estimates might well reveal data problems which could otherwise remain undetected and lead to erroneous econometric results. Furthermore, the index number approach can be applied to highly disaggregated data, whereas the econometric approach requires considerable aggregation. Finally, as Caves, Christensen, and Swanson have demonstrated, there may be applications in which it is attractive to combine the index number and econometric approaches.

In his final section Diewert observes that the existence of capital goods creates numerous problems for the practitioner. To his list of problems, I would like to add what I would consider to be the most serious practical problem: the economic entity may not be holding the stock of capital

goods which minimizes total cost for the *observed* levels of outputs and factor prices. For example, it is not uncommon for the analyst to find that yearly fluctuations in output and in the annualized cost of capital services are much larger than yearly fluctuations in capital. Estimation of productivity growth under such circumstances may require that the assumption of full equilibrium in each period be amended. Econometric approaches which assume that the firm is in equilibrium not unconditionally, but conditional on its level of capital stock, are now being developed and appear quite promising.

Interactions Between Industrialization and Exports

By HOLLIS B. CHENERY*

Sustained economic growth requires a transformation of the structure of production that is compatible with both the evolution of domestic demand and the opportunities for international trade. This transformation normally involves a substantial rise in the share of industry and—except for a few specialized mineral producers—a shift away from dependence on primary exports toward manufactured goods as a source of foreign exchange. There is considerable evidence that success in developing manufactured exports is critical to this process, and conversely that continued emphasis on import substitution will ultimately lead to a slowing down of growth.

Despite the amount of attention given to alternative strategies of trade and development since the work of I. M. D. Little, Tibor Scitovsky and M. F. G. Scott, there has been little attempt to examine the underlying relationships in quantitative terms. A fuller understanding of the various mechanisms that have been posited requires that the internal and external aspects of industrialization be examined together in a framework that brings out the several interactions among them.

This paper is drawn from a comparative study of sources of industrial growth in selected semi-industrial countries that have followed policies ranging from the extremes of export promotion to import substitution. The core of the analysis is a set of input-output accounts that permits changes in the structure of demand, trade, and production to be analyzed in comparable terms over periods of fifteen to twenty years. I will compare the main effects of different types of trade and development strategy on industrial growth and structure. The methodology emphasizes interrelations on the demand side, which tend to be neglected in other approaches. Attention is focused on the effects of early or late development of manufactured exports, which is a major source of the differences among strategies.

I. The Transformation of Production and Trade

The structural transformation of developing countries is characterized by a period in which the rising share of manufacturing in *GNP* approaches that of primary production and a significant portion of manufactured goods begins to be exported. Countries that have reached this stage have been alternatively described as "semi-industrial" or "newly industrialized."[1] Depending on the criteria used, there were between twenty and twenty-five such countries by 1970.

The present sample consists of seven of the sixteen principal semi-industrial countries identified by Bergsman: Korea, Taiwan, Colombia, Turkey, Yugoslavia, Mexico, and Israel (in ascending order of 1960 per capita income).[2] Japan and Norway, which had largely completed the transformation of production by 1960, are added for comparative purposes. The sample was selected primarily on the basis of the availability of input-output data covering fifteen years or more. Table 1 gives selected structural characteristics for the nine countries.

*World Bank.

[1]Joel Bergsman identifies sixteen significant semi-industrial countries, ten of which are the subject of the recent OECD study of "The Impact of the Newly Industrialising Countries."

[2]The remaining nine countries identified by Bergsman are: Egypt, the Philippines, Brazil, Portugal, Hong Kong, Singapore, Greece, Argentina and Spain. Marginal cases include India, Uruguay, Chile, South Africa, Thailand and Malaysia.

TABLE 1—INDICATORS OF STRUCTURE AND GROWTH

			Per Capita *GNP*		Share of *GDP*		
		Population (millions)	Level (US$1970)	Average Growth Rate[a]	Exports[a]	Manufactured Exports[a]	Value-Added in Industry[a,b]
Group A							
Korea	1955	22	131	–	1.6	0.2	13.1
	1963	27	149	1.6	4.9	1.2	16.9
	1973	33	323	8.0	31.8	24.3	29.7
Taiwan	1955	9	199	–	8.3	1.4	23.6
	1963	12	252	3.0	17.6	6.2	28.6
	1973	15	513	7.4	51.6	38.3	43.8
Israel	1955	2	950	–	11.5	4.9	31.6
	1963	2	1429	5.2	21.4	11.2	35.5
	1973	3	2374	5.2	28.3	15.1	36.7
Norway	1955	3	1244	–	40.7	10.4	35.2
	1963	4	2168	7.2	39.0	12.6	33.1
	1973	4	3179	3.9	43.4	19.1	30.4
Group B							
Yugoslavia	1955	18	329	–	6.6	2.0	41.7
	1963	19	510	5.6	15.6	11.8	40.8
	1973	21	813	4.8	22.3	12.5	41.4
Japan	1955	89	500	–	10.7	9.1	26.5
	1963	97	992	8.9	9.3	7.4	40.8
	1973	108	2349	9.0	10.3	8.8	42.5
Group C							
Colombia	1955	13	285	–	12.4	0.2	19.3
	1963	17	309	1.0	11.9	0.5	23.2
	1973	23	415	3.0	15.5	3.7	24.9
Turkey	1955	24	264	–	5.2	0.2	16.9
	1963	30	319	3.7	5.9	0.3	19.0
	1973	38	461	3.8	8.1	1.5	24.5
Mexico	1955	31	424	–	16.7	3.8	26.8
	1963	40	513	2.4	10.4	1.7	27.2
	1973	56	719	3.4	9.2	3.0	31.1

Source: See Kubo and Robinson.
[a]Shown in percent.
[b]Industry includes manufacturing and construction.

A. *The Sources of Growth*

An input-output model is used to provide a consistent framework for the analysis of growth and structural change. The same twenty-three-sector classification is used for each country, which leads to a comparable decomposition of output growth in each sector into the direct and indirect effects of increases in domestic demand, exports, and import substitution. For this purpose import substitution is defined for each sector by the reduction in the share of total supply that is provided by imports.

The model is based on the following accounting balances for each sector:

$$X_i = u_i(W_i + D_i) + E_i \tag{1}$$

$$M_i = m_i(W_i + D_i) \tag{2}$$

where X is total output, D is domestic final demand, W is intermediate demand, M is imports, E is exports, m_i is the share of imports in total supply, and u_i is the domestic share. Assuming that $W_i = \Sigma a_{ij} X_j$ the level of output can be expressed by the solution to the corresponding Leontief

model as

$$(3) \qquad X_i = \Sigma r_{ij}(u_j D_j + E_j)$$

The coefficients r_{ij} are the elements of the inverse of a Leontief domestic matrix in which the coefficients $(u_i a_{ij})$ represent the amount supplied from domestic sources.

Equation (3) makes it possible to solve for the increase in output of each sector, ΔX_i, in terms of increases in internal and external demand in all sectors (ΔD_j and ΔE_j) and changes in two sets of parameters (Δu_j and Δa_{ij}). The solution for ΔX_i can be expressed as the sum of four factors:[3]

(a) The expansion of domestic demand in all sectors (DD):

$$\sum_j r_{ij}^1 u_j^1 \Delta D_j$$

(b) The expansion of exports in all sectors (EE):

$$\sum_j r_{ij}^1 \Delta E_j$$

(c) Import substitution in all sectors (IS):

$$\sum_j r_{ij}^1 \Delta u_j (D_j^2 + W_j^2)$$

(d) Technological change (TC):

$$\sum_j r_{ij}^1 u_j^1 \sum_k \Delta a_{jk} X_k^2$$

The effects of trade policy are shown by terms (b) and (c), export expansion and import substitution. When there is no change in import proportions or in input-output coefficients, the last two terms vanish and sectoral growth is determined only by increases in internal and external demands.

B. *The Role of Trade*

Trade and development strategies are often characterized by a spectrum varying from inward to outward looking or from "import substituting" to "export led." The direct effects of these policy differences on production are shown most clearly by changes in the share of manufactured exports, which are given in Table 1. Since the sample illustrates a wide variety of development patterns, there is little difficulty in dividing the countries into three groups on this basis. *Group A*: Countries with high or rapidly rising manufactured exports: Korea, Taiwan, Israel, Norway. *Group B*: Intermediate cases: Yugoslavia and Japan. *Group C*: Countries with low manufactured exports: Colombia, Turkey and Mexico.

In the two intermediate cases, manufactured exports rose rapidly before 1960 but maintained a relatively high and stable share of *GDP* thereafter.

The four sources of growth of all manufacturing for these three groups of countries are given in Table 2. The subperiods are five to ten years, depending on the availability of input-output data. They extend from the mid-1950's to the early 1970's, except for Japan where it was possible to make approximate calculations for the prewar period. In all countries except Norway, the data cover part of the initial period of import substitution, which is particularly notable in the analysis of Korea, Taiwan, and Colombia. Thereafter the patterns diverge substantially.

In the four countries in Group A, the growth of manufacturing is increasingly due to the continued expansion of exports, which accounts for 50 percent or more of the total increase in output. In Korea and Taiwan, export expansion led to a rapid acceleration of industrial growth; but in Israel, Norway, and Yugoslavia the demand effects of export expansion were largely offset by import liberalization.

The countries in Group C are typical of a larger group (which includes India, Brazil, Chile, Uruguay, and Argentina) whose development strategy has been based on import substitution for several decades (see my book with Syrquin, Table 16). The decomposition of the sources of manufacturing growth shows that export expansion was the smallest of the four factors, accounting for less than 10 percent of the total increase.

[3]This formulation is discussed in the paper by Moises Syrquin and myself. The superscripts refer to time periods.

TABLE 2—SOURCES OF GROWTH IN MANUFACTURING OUTPUT

	Period	Average Annual Growth Rate	Percent of Total: Domestic Demand Expansion	Export Expansion	Import Substitution	Changes in Input-Output Coefficients
Group A						
Korea	1955–63	10.4	57	12	42	−11
	1963–70	18.9	70	30	0	0
	1970–73	23.8	39	62	−3	2
Taiwan	1956–61	11.2	35	28	25	12
	1961–66	16.6	49	44	2	5
	1966–71	21.1	35	57	4	4
Israel	1958–65	13.6	62	27	13	−2
	1965–72	11.3	71	49	−37	17
Norway	1953–61	5.0	65	36	−16	15
	1961–69	5.3	51	58	−19	10
Group B						
Yugoslavia	1962–66	16.6	74	25	−5	6
	1966–72	9.1	72	38	−22	12
Japan	1914–35	5.5	70	33	5	−8
	1935–55	2.8	71	−7	15	21
	1955–60	12.6	76	12	−3	15
	1960–65	10.8	82	22	0	−4
	1965–70	16.5	74	18	−1	9
Group C						
Colombia	1953–66	8.3	60	7	22	11
	1966–70	7.4	76	5	4	15
Turkey	1953–63	6.4	81	2	9	8
	1963–68	9.9	75	5	10	10
	1968–73	9.4	71	16	−2	15
Mexico	1950–60	7.0	72	3	11	14
	1960–70	8.6	86	4	11	−1
	1970–75	7.2	81	8	3	8

Source: See Table 1.

Colombia, the least industrialized of this group, illustrates the typical pattern of declining effects of import substitution with no offsetting rise in export effects. After import substitution is largely completed, manufacturing growth cannot exceed that of domestic demand and therefore tends to decline until there is a change in trade policy.

II. Trade-Development Sequences

The previous section established a rough grouping of countries based primarily on the role of manufactured exports in the structural transformation of the economy. I will now examine the differences in development-trade sequences at a less-aggregated level to ascertain the extent to which they vary among countries and industries. Before doing so, some of the differences in trade policies will be noted.

The seven developing countries all experienced some degree of balance-of-payments disequilibrium during the 1950's, reflected in foreign exchange shortages and quantitative restrictions (*QR*s), and exacerbated by overvalued exchange rates. The trade and exchange-rate regimes of four of these countries—Korea, Israel, Colombia and Turkey—have been compared to a larger sample by Anne Krueger. While Korea and Israel show progressive liberalization and reduction of *QR*s by the early 1960's, Colombia and Turkey maintained high levels of protection and import substituting policies for most of the period. Both the latter had intervals of liberalization and export expansion in the late 1960's.

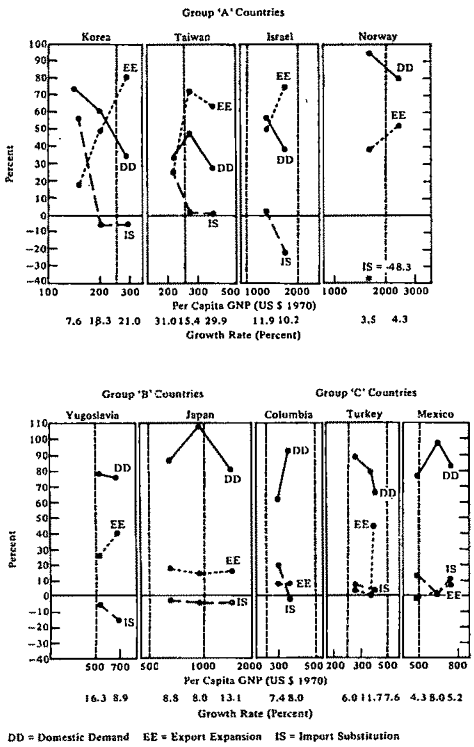


FIGURE 1. SOURCES OF INCREASE IN LIGHT INDUSTRY (PERCENT)

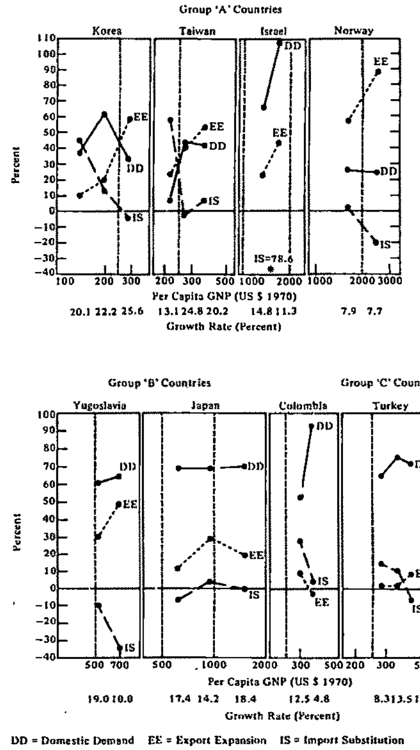


FIGURE 2. SOURCES OF INCREASE IN HEAVY INDUSTRY (PERCENT)

Yuji Kubo and Sherman Robinson have extended this comparison to the other countries in our sample. In Group A, Taiwan followed a sequence similar to Korea and liberalized trade even more fully by 1970. In Group C, Mexico followed a moderate form of import substitution strategy with relatively low levels of protection.

In summary, the trade policies of the countries in Group A (plus Japan) actively favored exports since the early 1960's, while those of Group C discriminated against them in varying degrees. In addition the transformation was affected by large inflows of foreign assistance to Korea, Taiwan, and Israel which made possible higher growth rates and more outward-looking trade and development policies.

To indicate the differences in trade-development sequences among sectors, the fourteen branches of manufacturing in our models have been aggregated into three groups: (a) light industry (food, textiles, clothing, wood products, etc.); (b) heavy industry (chemicals, metals, petroleum, etc.), and (c) machinery. Light industry includes sectors in which both demand and factor proportions favor early development while machinery typically develops at a relatively late stage. The three principal sources of growth, expressed as percentages of each sector's increase in output (as in Table 2), are shown in Figures 1, 2, and 3 for each of these sectors.

Outward-Looking Sequences. In Group A countries, each sector shows the same decline in import substitution and rapid rise of exports as a source of growth that was indicated in Table 2. This shift is earlier and more pronounced in light industry, where import substitution is only significant in the first period, and takes place last in machinery. These differences also persist at less-aggregated levels. There is a corresponding change in the pattern of exports in

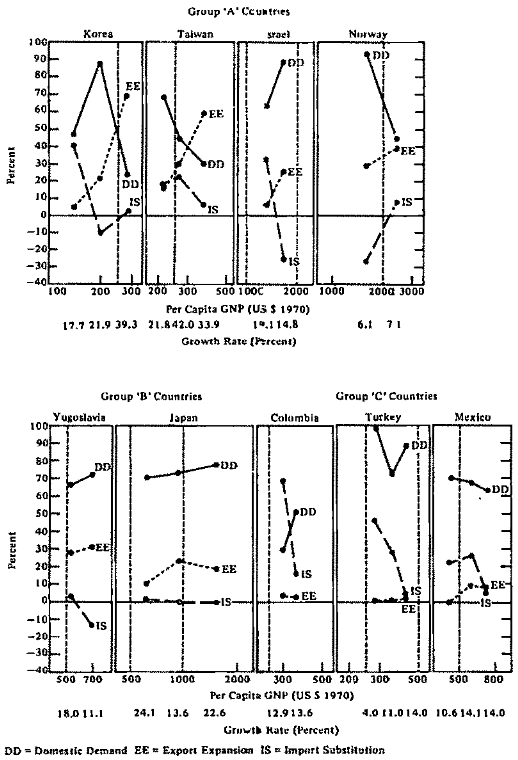


FIGURE 3. SOURCES OF INCREASE IN MACHINERY (PERCENT)

each country (not shown) with a growing component of heavy industrial products.

Despite these differences in degree, the major impression from these comparisons is a general similarity in the trade-development sequences of each sector in the three developing countries in Group A. After an initial period of strong import substitution, export expansion became the major source of industrial growth in Korea and Taiwan in each major sector, and also led to an acceleration of growth in each. In Israel, however, exports predominated only in light industry.

A final phase of import liberalization in each sector is shown by Norway, Israel, and Yugoslavia. In this phase export expansion is partly offset by increased imports, so that the rate of sectoral growth is determined primarily by domestic demand.

Inward-Looking Sequences. Although there are some significant differences in timing, the inward-looking countries of Group C indicate the effects of the exhaustion of import substitution possibilities in all sectors. The failure to develop manufactured exports (except on a modest scale in light industry) has led to the decline of the rate of growth shown in both light and heavy industry, but not yet in machinery.

Even though Colombia, Turkey, and Mexico are relatively large countries and have had fairly rapid rates of growth, the expansion of the domestic market has not offset this failure. A more detailed analysis shows these countries lagging particularly in machinery and metal products, sectors in which the countries in Group A have had above average growth.

In summary, the general features of the inward-looking pattern carry over to each of the major sectors, as in the case of the export-led strategy. While the opportunities for import substitution persist longer in heavy industry and machinery than in light industry, its ultimate decline is similar.

III. Concluding Remarks

This paper illustrates an approach to the analysis of structural change in which internal and external factors are treated together in an interindustry framework. The method is adapted from techniques used in development planning and takes advantage of information collected for this purpose. It can be extended to encompass production functions and factor use by sector in order to provide a more complete analysis of the sources of growth and structural change.

The effects of trade policy on industrial structures that are revealed by this analysis are quite striking. Although import substitution is an important feature of early stages of industrialization in all developing countries, it can be accelerated or retarded by trade policy. The later stage of expansion of manufactured exports is more susceptible to policy influence and is shown to have a large effect on the subsequent course of industrial development.

This methodology can also be used to explore the effects of the balance-of-payments constraint on the pattern of develop-

ment. In this context the development of manufactured exports appears even more important as a source of foreign exchange than as a source of demand because it provides one of the principal means of exploiting comparative advantage and of avoiding balance-of-payments bottlenecks.

To complete the linkage between industrialization and export growth, it would be necessary to examine the changes in comparative advantage that result from the acceleration of growth and learning by doing in successful export-led strategies. This process, which is explored in my paper with Donald Keesing, has been a major factor behind the growing share of the semi-industrial countries in world markets for manufactures. Their success in exporting manufactured goods has in turn contributed to more rapid industrial growth in a cumulative process.

REFERENCES

J. Bergsman, "Growth and Equity in Semi-Industrialized Countries," staff work. paper no. 351, World Bank 1979.

H. B. Chenery and D. B. Keesing, "The Changing Composition of Developing Country Exports," staff work. paper no. 314, World Bank 1979.

Hollis B. Chenery and Moises Syrquin, *Patterns of Development 1950–1970*, London 1975.

______ **and** ______, "A Comparative Analysis of Industrial Growth," in R. C. O. Matthews, ed., *Measurement, History and Factors of Economic Growth*, New York 1979.

Anne O. Krueger, *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*, New York 1978.

Y. Kubo and S. Robinson, "Sources of Industrial Growth and Structural Change: A Comparative Analysis of Eight Countries," mimeo., World Bank 1979.

I. M. D. Little, Tibor Scitovsky, and M. F. G. Scott, *Industry and Trade in Some Developing Countries: A Comparative Study*, Oxford 1970.

Trade Policy as an Input to Development

By ANNE O. KRUEGER*

My topic is the question: what difference does the set of commercial policies chosen by a developing country make to its rate of economic growth? Three points are salient. First, in its present state, trade theory provides little guidance as to the role of trade policy and trade strategy in promoting growth. Second, the empirical evidence overwhelmingly indicates that there are important links between them. Third, a number of hypotheses as to the reasons for these links have been put forward, but there is not as yet sufficient evidence to enable us to estimate their relative importance.

Turning first to theory, there are many static propositions but few useful theorems about the effects of alternative trade policies on growth. Clearly there are gains to be achieved through trade in the development process. Even the trade and growth models along Corden-Johnson lines are based upon differential rates of change in capital-labor ratios in two-country, two-commodity worlds under assumptions of free trade. They provide little indication of the quantitative importance of trade as a contributor to growth, and still less insight into the probable orders of magnitude of the losses in attainable growth rates that may be incurred with departures from free trade.

To be sure, once the assumption that there are only two goods is abandoned, theory suggests that activity in production of tradables should be undertaken to the point where the international marginal rate of transformation (*IMRT*) equals the domestic marginal rate of transformation (*DMRT*), with no production in lines where domestic opportunity cost exceeds the international price ratio. An allocation of resources satisfying this criterion would be optimal in the absence of any dynamic considerations.

Theory does not, however, indicate how many activities are likely to be undertaken. Nor does it suggest the relative importance of exporting and import-competing activities in an optimum allocation, or how that allocation would change with growth. Worse yet, there is nothing in theory to indicate why a deviation from the optimum should affect the rate of economic growth. Most growth models suggest that there are once-and-for-all losses arising from nonoptimal policies with lower levels of income resulting from them but no change in growth rates.

Turning from theory to practice, developing countries' trade policies have fallen into two distinct categories. One group of developing countries has adopted trade policies which diverge from the optimality criterion, often by a large amount, by protecting their domestic industries. These "import substitution" policies have been employed to stimulate domestic production on the theory that nonagricultural sectors must grow at a rate above the rate of growth of domestic demand, and can do so only insofar as additional production substitutes for imports. The other category, "export promotion," has consisted of encouragement to exports, usually beyond the extent that would conform to the *IMRT*=*DMRT* criterion. Countries adopting an export-oriented trade strategy have generally experienced rapid growth of traditional exports, but even more rapid growth of nontraditional exports.

Experience has been that growth performance has been more satisfactory under export promotion strategies (meant as a general bias toward exports and not as a package of specific measures to encourage selective exports of particular items themselves induced by a bias toward import substitution) than under import-substitution strategies. While it is impossible to specify a

*University of Minnesota. I am indebted to James M. Henderson for helpful suggestions on a first draft of this paper. The research underlying the paper was supported in part by the National Science Foundation, under grant No. NSF/SOC 77-25776.

particular model of the growth process that will simultaneously satisfy all observers, the relationship between export performance and growth is sufficiently strong that it seems to bear up under many different specifications of the relationship. It has been tested over many countries for: 1) rates of growth of real *GNP* and of exports (see Michael Michaely); 2) for real *GNP* net of exports and exports (see Bela Balassa); and 3) for rates of growth of *GNP* as a function of rate of capital formation, aid receipts, and export growth (see Constantine Michalopoulos and Keith Jay). Time-series and cross-section data have been pooled, so that deviations of countries' growth rates from their trends have been estimated as a function of the growth of export earnings (see my book, p. 271ff). In all of these specifications, rate of growth of exports has turned out to be a highly significant variable. While the "success stories" of Korea, Taiwan, and Brazil are well known, there are enough other observations, both for different time periods in the same country (as for example Turkey and the Philippines) and of countries (including on the positive side Ivory Coast, Colombia, and Malaysia and on the negative side India, Argentina, and Egypt), so that there is little doubt about the link between export performance and growth rates.

Moreover, it seems clear that export performance is a function in large part of governmental policies. While an export promotion strategy will not always be successful in generating more export growth (especially if policies affecting the domestic market are inappropriate), certainly policies adopted to encourage import substitution, especially when they include overvalued exchange rates and quantitative restrictions upon imports, retard the growth of exports. Moreover, in instances where exports have fortuitously risen (as, for example, through favorable terms-of-trade changes as in Chile in the mid-1960's) or where other sources of foreign exchange such as foreign aid have grown (as in Egypt), economic performance has not matched that attained when exports have grown rapidly.

The central question, then, is why such a difference in growth performance should be associated with export promotion contrasted with import substitution. There are three major hypotheses, and each undoubtedly contains some explanatory power. Their relative importance probably varies from country to country, and of course they are not entirely independent. Elaboration of these hypotheses provides an indication of the many ways in which the choice of trade strategy and its implementation can significantly affect the rate of economic growth. In view of the fact that empirical testing of alternative hypotheses has not yet been undertaken, there can be legitimate differences in opinion.

The first hypothesis is that technological-economic factors imply an overwhelming superiority for development through export promotion. These factors include such phenomena as minimum efficient size of plant, increasing returns to scale, indivisibilities in the production process, and the necessity for competition. According to this hypothesis, failure to take advantage of the opportunities to exploit these phenomena through trade significantly impairs the attainable rate of growth. A second hypothesis is that differences in growth rates have resulted, not from the choice of trade strategy per se, but rather from excesses in the ways in which import substitution policies were administered. The third hypothesis is that policies adopted in pursuit of an export promotion strategy are generally far closer to an optimum, both in the $DMRT = IMRT$ sense and with respect to the domestic market, than are those adopted under import substitution. Under this interpretation, the role of trade policy is to constrain policymakers in such a way that they do not impede the growth rate as much as they otherwise would.

Both the first and second hypotheses are consistent with the notion that the nonagricultural sector of most developing countries is, in some sense, an "infant industry," and requires some stimulus for growth. The third, by contrast, essentially takes the negative view, that markets would function well

and provide satisfactory growth if only policymakers would abstain from counterproductive intervention.

The first hypothesis really amounts to an assertion that the gains from trade, especially for developing countries, are so sizable that the losses associated with import substitution significantly reduce the rate of return on factor accumulation. On the negative side, domestic markets are extremely small in most developing countries, and attempts to replace imports result in the construction of plants of less-than-efficient minimum size, while simultaneously generating an oligopolistic or monopolistic market structure. As import substitution proceeds, new activities are increasingly capital intensive and inefficiencies from below-minimum-efficient size increase. On the positive side, so the argument runs, export promotion permits entrepreneurs to base their plans on whatever size plant seems appropriate: size of domestic market is no longer a virtually binding constraint, as it is when the activity is profitable only because of very high rates of effective protection. Moreover, monopoly positions arise less frequently under export promotion, as exporters face competition from abroad as well as from other domestic producers.

Export promotion may also be more efficient in permitting rapid expansion of profitable activities; by contrast, under import substitution, most activities are constrained to expanding at approximately the same rate: inefficient firms and sectors expand approximately as rapidly as efficient ones. In this view, potential export lines consist of a number of industrial products (girls' sneakers, wigs, tennis rackets, engine parts, plywood, and so on) and it is as much a matter of the right entrepreneur, and the right specialized product, as choosing the "right industry" that is necessary for rapid growth. To be sure, factor proportions and comparative advantage may result in greater profitability of relatively labor-using industries, but the basic notion is that there are thousands of industrial products, and that, among relatively labor-intensive activities, the ones which will develop into exports will be those in which there are firms with good management and an ability to utilize factors of production efficiently.

A final aspect of the technology-related view of the advantages of export promotion has to do with factor proportions. Given the vast disparity in capital-labor ratios of the industrial sectors of the developed and developing countries, the opportunity for trade represents a means for shifting the demand for labor outward more rapidly than the import-substitution strategy permits. If there are differences of, say, two-to-one and six-to-one in capital-labor ratios between activities at the prevailing wage-rental ratio, while the rate of capital accumulation is the binding constraint on expansion of employment in the urban sector, an allocation of additional capital to the labor-intensive activity for export will permit an upward shift in the demand for labor three times as great as that which would occur if import substitution dictates the start of the more capital-intensive activity. To be sure, the expectation is that the more rapid rate of growth of demand for industrial labor would drive up the urban wage once the demand for labor was rising more rapidly than the labor force, but this is precisely a desired outcome of policy.

Even within the "technology" view of the superiority of the export-oriented policy, it is important to learn the extent to which gains accrue because of indivisibilities, competition, and minimum-efficient-size considerations, and the extent to which they are the result of the ability to trade "surplus" factors. For, if the first set of factors is of preponderant importance, regional trading arrangements among *LDC*s may offer some hope for increasing gains from trade. However, if gains result primarily from "factor proportions" trade, the scope for intra-*LDC* trade as a substitute for *LDC-DC* trade is far more limited. There is already evidence that "factor-proportions" trade is important (see my coauthored book), although its relative weight, contrasted with other technology considerations, is not known.

The second hypothesis focuses upon the costs of import substitution policies as in

fact carried out, and suggests that alternative means of achieving import substitution might have avoided them. According to this view, the failure of import substitution resulted from the excesses of the particular ways in which domestic industries were encouraged: extreme currency overvaluation combined with quantitative restrictions provided the equivalent of prohibitive tariff protection; techniques of allocating import licenses were employed which prevented competition among domestic firms and rewarded entrepreneurs for license-getting abilities rather than their cost-minimizing performance; and excessive and detailed quantitative controls were employed over many aspects of economic activity. One of the costs was the failure of export earnings to grow as much as they would have under "better" import-substitution policies; that in turn led to "stop-go" patterns with their attendant costs. Simultaneously, the emerging "foreign exchange bottleneck" had both direct and indirect impacts upon the structure and growth of the economy. In particular, efforts at "import substitution" stopped being geared toward development of economic new industries, and became focused upon "foreign exchange saving," often in highly irrational and indiscriminate ways, which further distorted the system.

The third view denies the need for any bias toward exports and implicitly or explicitly asserts that growth would be optimal in the absence of intervention. A bias toward exports is therefore better than one toward import-substitutes only because policies are less distortive. In this view, an export-oriented strategy imposes constraints on policymakers, both in what they can attempt to do, and in making them aware of the costs of mistakes. Policymakers receive feedback in a relatively short time period as to the costs of their policies. Also, it is infeasible to rely upon quantitative controls: the international price, at least, cannot be administered and to that extent, more generalized forms of incentive, including a relatively realistic exchange rate, must be employed. Indeed, it is argued that incentives cannot be as biased toward export promotion as they can be toward import substitution. This is precisely because to do so would require either export subsidization (whose costs would be immediately evident through the drain on the budget) or such a degree of currency undervaluation that a current account surplus would absorb much of the country's savings potential.

According to this third line of argument, constraints upon policymakers go well beyond the inability to impose too great a bias toward exports. For example, it is virtually impossible to administer any highly protective system for intermediate and capital goods imports if exporters are expected to compete in international markets: they must be permitted ready access to imported raw materials, intermediate goods, and capital equipment. To impose any comprehensive system of licensing or controls would entail delays and other costs, inconsistent with the export strategy. Thus, the commitment to an export-oriented development strategy implies a fairly liberal and efficient trade regime, and thus prevents paperwork, delays, bureaucratic regulation, and other costs that can arise under import substitution. This in turn limits the restrictions that can be imposed on capital account. More generally, under an export promotion strategy, there is an international market in the background: it functions as a constraint upon economic behavior, both of entrepreneurs and of government officials, and simultaneously provides feedback to them as to the success or failure of policies in terms of their objectives.

Undoubtedly, all three approaches to the differential in economic performance contain elements of truth. There are export opportunities that are passed up under import substitution where indivisibilities or increasing returns within a range would permit sizable gains in output. There are also high-cost import substitution activities which, if never undertaken, would have freed resources for considerably more productive use, even within an import-substitution strategy. Likewise, the international market has served to constrain policymakers and induce them to abandon uneconomic

policies sooner than they otherwise would have done. Knowledge is not yet far enough advanced to determine the relative importance of the alternatives. It will not be until we have far more information than is currently available about the order of magnitude of indivisibilities and minimum size plant contrasted with size of markets in *LDC*s, and also about the determinants of politicians' and bureaucrats' behavior. Moreover, it is certain that the primary sources of growth are internal, and that there is no magic formula, or single policy change, that can by itself account meaningfully for differences in economic performance.

Nonetheless, experience has clearly demonstrated the importance of access to international markets in providing a means of permitting more rapid growth than would otherwise be feasible. Given the enormous difficulties and costs of achieving the institutional and other changes that economic growth requires, it is probable that trade policy changes have a higher rate of return to *LDC*s than most other feasible policy changes. It is, of course, to be hoped that protectionist pressures in the developed countries do not result in fewer opportunities for the *LDC*s. If such protectionist measures are taken, they will lower the rate of return to outward-oriented trade strategies. They will however, for the foreseeable future, still leave that rate distinctly above the returns from a policy of persisting with inward-oriented growth.

REFERENCES

B. Balassa, "Exports and Economic Growth: Further Evidence," *J. Develop. Econ.*, June 1978, *5*, 181–89.

W. M. Corden, "The Effects of Trade on the Rate of Growth," in Jagdish Bhagwati et al., eds., *Trade, Balance of Payments, and Growth*, Amsterdam 1971.

H. G. Johnson, "The Theory of Trade and Growth: A Diagrammatic Analysis," in Jagdish Bhagwati et al., eds., *Trade, Balance of Payments, and Growth*, Amsterdam 1971.

I. Kravis, "Trade as a Handmaiden of Growth: Similarities between the Nineteenth and Twentieth Centuries," *Econ. J.*, Dec. 1970, *80*, 850–72.

Anne O. Krueger, *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*, Cambridge, Mass. 1978.

______ **et al.,** *Alternative Trade Strategies and Employment*, Chicago, forthcoming.

M. Michaely, "Exports and Growth: An Empirical Investigation," *J. Develop. Econ.*, Mar. 1977, *4*, 49–53.

C. Michalopoulos and K. Jay, "Growth of Exports and Income in the Developing World: A Neoclassical View," disc. paper no. 28, A.I.D., Nov. 1975.

Government Policies and Changing Shares in World Trade

By CHARLES P. KINDLEBERGER*

Many factors, including differential rates of economic growth, the product cycle, different distances of various countries from markets, and the decline in transport costs of goods at a faster rate than their production costs, play a part in changing countries' shares in world trade. The question is whether in addition to these factors, governmental policy plays a role. There is evidence on the one hand that the metropolitan country has in the past dominated the imports of its colonies, to varying degrees, and that the Ministry of Trade and Industry (*MITI*) in Japan plays a powerful role in keeping out imports and promoting Japanese exports; and on the other hand that policies of export promotion, parastatal credit facilities, tax advantages and the like are ineffectual because of either retaliation by competitors, or low elasticities of export supply. In between there is room for a wide range of compromise positions suggesting that government policy has some, but not an overwhelming, effect on competitive export shares. I restrict my discussion to the policies of the exporting countries, since discriminatory import prohibitions clearly have effects (aside from smuggling). Moreover, the discussion is limited to policies designed to affect the competitive position in exports, and neglects indirect effects such as subsidies for *R&D*, military research in aircraft with spillovers into civil aviation, and taxation of corporate income from abroad that affects the decision to invest abroad or export. The treatment is historical to a considerable extent and confined to leading industrial countries. The tentative overall conclusion is that, except when a country can dictate the trade policy of the importing country, government policies have a negligible effect on the competitive share of country's exports. It would extend the range of the discussion too widely to include economic welfare through blockade, sanctions, licensing of strategic goods, but I offer the *obiter dictum* that the efficiency of such measures, when they do not boomerang, is highly questionable. On occasion to be sure, policies designed for separate purposes will have unintended effects on export shares, as for example, the effect of German reparations in kind on the market for British coal.

In 1913, France provided 82.5 percent of Algerian imports, Britain 64 percent of those of British India but only 14.2 percent of imports of British Malay States, the United States 50 percent of those of the Philippines, the Netherlands 33.3 percent of those of the Netherlands East Indies (see Paul Bairoch, p. 45). High proportions—the highest being 86 percent for Madagascar for France—reflect active trade policies. With independence, these proportions changed somewhat but by no means completely. In 1958, for example, Nigerian imports were 44 percent from the United Kingdom and 2 percent from France; while those of French West Africa, not far away, were 65 percent from France and 2 percent from Britain.

Where a country has influence but not dominance, it may try to lower tariffs abroad to enlarge imports as a whole, and presumably its share. It is noteworthy that Britain sought, as a condition for its last subsidies to its Allies in the Napoleonic Wars, to urge them to lower tariffs on goods exported by Britain, but with no success (see John Sherwig, p. 311). In World War II and its aftermath, the United States collected signatures to statements favoring low tariffs in the Atlantic Charter, Article VII of the Lend-Lease settlement, the Anglo-American Financial Agreement (British loan), and the European Recovery Program bilateral

*Massachusetts Institute of Technology.

agreements, but then did not object when various countries discriminated against *U.S.* exports until the reestablishment of convertibility in 1958.

In the heyday of British free trade, the Board of Trade believed that unilateral tariff reduction by the United Kingdom would enlarge the British share of exports, through some such mechanism as Hume's Law, as a result of which imports create exports. Repeal of the Corn Laws was sought 1) as an answer to the Zollverein; and 2) to prevent Germany from industrializing by providing her a renewed comparative advantage in exporting grain (see John Bowring, p. 55). The political economists of the Board of Trade, however, were not above lobbying directly in favor of tariff reductions abroad (see Barrie Ratcliffe, pp. 67ff). The name given to these policies is free-trade imperialism, which describes how Britain through advocating and engaging in free trade, tried to prevent or at least slow the rise of competitive production abroad (see John Gallagher and Ronald Robinson; Bernard Semmel). Note that the free-trade imperialist is thought by other countries to be working for its own interest, and that of its abundant factor in the Stolper-Samuelson analysis, whereas it is actually persuaded that free trade is a public good that benefits all.

The case that troubles the United States today is Japan, where commitments have been frequently given to the policy of opening up Japanese markets to import competition, although the process seems slow in practice. As the Japanese export surplus mounted, *U.S.* trade negotiators have insisted with increasing vehemence that it was time for the country to make good on its commitments and to reciprocate on the trade concessions made to it in admitting its textiles originally and later its steel, television sets, cameras, ships, and automobiles. Japan is happy to buy raw materials and foodstuffs from the United States that are in short supply, especially coal, oil, soya beans, and steel scrap. However, when a domestic interest however small, such as citrus fruit, is threatened by imports, Japan finds it difficult to reduce import restrictions.

An important element of free-trade imperialism for a time was the prohibition of exports of machinery or such critical inputs to foreign production as coal, wool, raw silk, etc. These policies were difficult to sustain because of smuggling, objection by local producers calling for *laissez-faire* to be applied to them, and because they clearly stimulated production abroad as a foreign tariff would have done, by raising the price (see A. E. Musson). It is noteworthy that the Nixon Administration briefly contemplated prohibiting the export of technology through licensing and transfer through multinational corporations, but decided ultimately that it was impossible to prevent technological diffusion.

Reciprocal tariff reductions evidently affect competitive export shares, even when they are worldwide, since not all supply elasticities are identical. The effects of discriminatory reductions are increased in a customs union or free-trade area. The Eden treaty of 1786 lasted only three years before abandonment in the French revolution, but this was thought to have given one-sided advantages to Britain and hurt industrialization in France. The Zollverein discriminated against British coarse textiles, and later cotton yarn. The Cobden-Chevalier treaty of 1860 ushered in a wave of reciprocal reductions among countries in Europe that probably raised the ratio of trade to domestic production but left competitive shares not much altered. Italy is said to have developed the technique of trade treaties to a fine point after 1878 by raising tariffs generally and then reducing them in trade treaties to gain outlets, especially for citrus fruits and wines (see Frank Coppa). A recent study blames Gladstone for not applying "negotiating tariffs" that could be reduced in reciprocal agreements, and blames much of Britain's loss of its markets in Europe after the 1880's on this omission (see John Gaston). It is hard to know how much stock to put in this argument. Tariff increases on the Continent and in the United States did fence off some British goods, however, it is not clear whether British business and political leaders were ready to abandon free trade, even if Gladstone had sought to do

so. It is also not certain whether imposition of "fighting tariffs" would have led to Continental tariff reductions rather than to retaliation. Retaliation was not unknown in tariff wars between France and Italy, and Germany and Russia. It became the rule in the 1920's and 1930's, and tariffs, imposed to acquire a bargaining counter to be traded away, often were found still in place and limiting trade.

British and Dutch capacity to enforce the nonbinding agreement to reduce tariffs reached at the World Economic Conference of 1927 was limited by the fact that as free-trading countries they had no step that they could take conditional on the action of others. Retaliation and threats to retaliate on the part of more than thirty countries, moreover, did not prevent President Hoover from signing the Hawley-Smoot tariff of 1930, which reduced total trade and inevitably shifted shares.

The outstanding government policy aimed at sustaining exports and enlarging a country's competitive position was probably Empire preference, finally negotiated by the British Commonwealth at Ottawa in the summer of 1932. Pressure for Empire preference had been exerted by Joseph Chamberlain in the 1890's, and during the 1920's by Conservatives of the L. S. Amery variety, who favored a cooperative Empire arrangement in lending, migration, and trading to discriminate against the outside world. Standard British exports kept out of the Continent and the United States by tariffs and rapid economic growth in import-competing products would be compensated for in part by enlarged exports to the Commonwealth (before 1913, the Empire). The share of British exports going to the Empire declined from roughly 35 percent in the late 1850's to 25 percent in the early 1870's before reaching a peak of 38.5 in 1902 and settling again at 35 percent in 1913. This result was achieved with no policy action by Britain, though Canada unilaterally applied discrimination in favor of British goods in 1900 and elicited retaliation against Britain from Germany which denounced the 1865 treaty. By 1928 the share going to the Empire had risen to 45 percent, but subsequent Ottawa agreements failed to raise it higher. In the postwar period as Commonwealth preferences were being dismantled under vigorous attack from the United States, the share of British exports to the Commonwealth rose to 53 percent in 1948 and remained almost that high at 50 percent in 1958. Government policy in Britain had little to do with the competitive position of Britain outside or inside the Commonwealth. Ian Drummond observes that the Empire worried more about tariffs than any other economic issue, though they dealt with economically insignificant goods, and should have spent the time dealing with monetary, and especially exchange rate policy (pp. 214–16).

Where reciprocal tariff concessions did make a difference in market shares was in the European Common Market and the European Free-Trade Community. Common Market exports rose from 31 percent of world trade in 1953 to 38.3 percent in 1961 and 39.7 percent in 1967, with trade in manufactures for the same period rising from 34.1 to 41.8 to 42.2 percent (see Charles Jeanneret-Grosjean, pp. 266–69). Intracommunity trade rose from 28.8 percent of total trade of the six in 1953 to 43.7 percent in 1967 (for manufactures from 24.2 percent to 40.7). How much of this was due to the Common Market and how much to faster growth on the part of Common Market countries compared to the rest of the world, I cannot determine. It should be noted, however, that the gains from this exchange, which were largely of manufactured goods for manufactured goods, may have been less than slower-growing trade with *LDC*s based on wider comparative advantages. European Free-Trade Area trade declined from 26.1 percent of the world total in 1953 to 20.1 percent in 1966, with British trade accounting for most of the decline—15.3 percent in 1953 and 10.5 percent in 1966. Part of the latter decline may have reflected policy in small part, as Britain rationalized the cotton textile industry by reducing capacity. For the most part, however, this and presumably other declines in exports were the result of the product cycles: British exports of cotton

textiles indicate the product cycle under first growth and then aging. Britain accounted for something close to 100 percent of exports of cotton textiles at the beginning of the nineteenth century. At the end, it was responsible for 42 percent of the world market in all textiles and clothing. By 1929 this share had fallen to 33 percent, by 1950 to 27 percent, and by 1959 to 15 percent (see Alfred Maizels, p. 363). These figures are sustained by woolens and synthetic fibers. In cotton yarns and piece goods, the share is far lower.

II

Financial policies conceivably affect the export competitive position of countries, but strong counterexamples are readily cited. In 1913 France lent heavily to Russia, Britain and Germany did not, but the respective shares of Russian import trade were 4.2, 12.6, and 47.5 percent (see Bairoch, p. 43). Beginning in the 1930's, various countries assisted their exporters by government intermediate credits; ultimately all did so through the Export-Import Bank in the United States, the Export Credit Guarantees Department in the United Kingdom, also COFACE in France, and similar institutions elsewhere. Total trade may have been enlarged but competitive shares were only marginally affected by the differences in terms for most buyers. Such would not be true where political factors entered, for example, in the substantial West German credits made available for exports to the Peoples Democratic Republic and other Eastern socialist countries.

Tied aid would of course affect competitive shares in a world of partial equilibrium, where everything else was unchanged. In the general equilibrium conditions of the real world, one cannot be certain that tied aid adds to a country's export capacity, since it may be used to obtain goods that would otherwise be bought commercially, and the foreign exchange thus saved may be diverted to other sources of supply.

Taxation works the same way. Starting from scratch, it is evident that Ireland, for example, can entice multinational corporations to establish plants through subsidies to plant construction, and expand exports of manufactures by long holidays on income taxes on profits from exports. In other cases with on-going trade, one cannot be so sure. The change from levying indirect taxes on production to levying them on consumption that accompanied the adoption of the value-added tax (*VAT*) by the Common Market, with border adjustments calling for adding *VAT* to imports and subtracting it from exports, was attacked as a subsidy to exports on the one hand, and rationalized by the conclusion that the exchange rate would adjust over time, on the other. In any event, it seems unlikely that the underlying theory that indirect taxes are all passed forward, instead of falling to some extent on inputs or producers profits can be justified. To the extent that indirect taxes are not passed forward, border adjustment is a subsidy to exports.

Financial policy affects exports indirectly through the multinational corporation by influencing its choice whether to export to a market or produce in it. It seems likely that the product cycle and technological diffusion are the major basis for choice, but such factors as the insurance available through the Overseas Private Investment Corporation, the tax credit on foreign-earned income, and the definition of where income is earned for tax purposes, may affect the decision at the margin. Critical to the effect on the export position, of course, is the counterfactual issue whether direct investment substitutes for exports, or takes place at a time when exports are declining inexorably (see Gary Hufbauer and F. Michael Adler).

The same inability to decide the extent to which government affects a country's export position extends to foreign exchange policy. Those who believe that currencies can become over- or undervalued accept that governmental exchange policies can affect shares in world trade; purchasing power parity believers hold that prices adjust to exchange rates rapidly so as to moderate and even to eliminate this source of financial incentive or penalty. It is surprising to most of us that Germany and Japan have maintained their shares of world exports despite substantial appreciation, and that

the depreciation of the dollar has not done more to reverse the decline in the United States share in world manufactures from its 26.6 percent in 1950, to 18.5 percent in 1970 and 16.5 in 1977.

Economists with a monetary approach to the balance of payments might maintain that reductions in the money supply enlarge a country's share in world trade as those with an unsatisfied demand for cash balances expand exports in order to build them to the desired level. I am sceptical of this. The normal way to obtain cash is through borrowing. But under foreign exchange control and bilateral clearing, one crucial monetary decision affects a country's share in world exports. I refer to the central bank choice between the waiting and the payments principle analyzed by P. Nyboe Andersen. Under the waiting principle, exporters are not paid local money until imports pay local funds into the clearing account. Under the payments principle, the central bank pays out to exporters on the basis of funds accumulated in clearing accounts. The waiting principle discourages exports, the payments principle encourages them. But the point is relevant only to the 1930's.

III

Trade and financial policies appear to have the major effect among policies on competitive shares, and that effect is not highly significant. I think it would be extraordinarily difficult to find a trace in the statistics of export promotion schemes—trade missions, consular representation, speeches by high officials urging the merits of overseas sales. It may be that the Ministry of Trade and Industry working closely with Japanese industry has produced a substantial increase in exports through consciousness raising, different from financial incentives.

That consciousness of the opportunities available in exporting is important cannot be denied. E. E. Williams' *Made in Germany*, written in the 1890's, compared the German push for exports with the more languid British attitude. After World War II, it was said that the more vigorous German businessmen chose export jobs because they felt hemmed in by fifteen years of autarky and war and wanted to see the world. While such attitudes exist, they may not be subject to manipulation.

I conclude that it is possible for national policies to restrict trade, as in autarky and in discriminatory tariffs, often historically imposed on dependencies by the metropole, but that it is hard for policy to push exports. Under some circumstances Hume's Law may enlarge the market for exports by admitting more imports when the exports in question are in demand and restricted mainly by availability of foreign exchange. It may be possible for short periods to expand export shares through financial incentives. In the long run, however, national export shares are determined by deep-seated changes in comparative advantage, originating in the forces that make for more rapid or slower growth. Policy can do little to mold these.

REFERENCES

P. Nyboe Andersen, *Bilateral Exchange Clearing Policy*, Copenhagen 1946.

P. Bairoch, "La place de la France sur les marchés internationaux," in Maurice Lévy-Leboyer, ed., *La position internationale de la France*, Paris 1977.

J. Bowring, "Report on the Prussian Commercial Union," in *Parliamentary Papers*, Vol. 21, London 1840.

F. J. Coppa, "The Italian Tariff and the Conflict between Agriculture and Industry: The Commercial Policy of Liberal Italy, 1860–1922," *J. Econ. Hist.*, Dec. 1970, *30*, 742–69.

Ian Drummond, *British Economic Policy and the Empire, 1919–1939*, London 1972.

J. Gallagher and R. E. Robinson, 'The Imperialism of Free Trade," *Econ. Hist. Rev.*, No. 1, 1953, *6*, 1–15.

J. W. T. Gaston, "Policy Making and Free-Trade Diplomacy: Britain's Commercial Relations with Western Europe, 1869–1886," unpublished doctoral dissertation, Univ. Saskatchewan 1975.

G. C. Hufbauer and F. M. Adler, "Overseas Manufacturing Investment and the Balance of Payments," Tax Policy Research Study No. 1, U.S. Treasury Department, Washington 1968.

Charles A. Jeanneret-Grosjean, *Les effets de structure et les effets de concurrence dans les échanges exterieurs des pays industrialisés, entre 1953 et 1966*, Berne 1971.

Alfred Maizels, *Industrial Growth and World Trade, An Empirical Study of Trends in Production, Consumption and Trade in Manufactures from 1899–1959 with a Discussion of Probable Future Trends*, Cambridge 1963.

A. E. Musson, "The 'Manchester School' and the Export of Machinery," *Bus. Hist.*, Jan. 1972. *14*. 17–50.

B. M. Ratcliffe, "The Reform Campaign in France, 1831–36," *J. Euro. Econ. Hist.*, Spring 1978, *7*, 61–138.

Bernard Semmel, *The Rise of Free Trade Imperialism: Classical Political Economy, the Empire of Free Trade, and Imperialism, 1750–1850*, Cambridge 1970.

John M. Sherwig, *Guineas and Gunpowder, British Foreign Aid in the Wars with France, 1793–1815*, Cambridge, Mass. 1969.

E. E. Williams, *Made in Germany*, 2d ed., London 1896.

DISCUSSION

CARLOS F. DIAZ ALEJANDRO, Yale University: Since Adam Smith, the professional mainstream has felt that the closer commercial policies are to free trade the higher the growth rate will be. Yet, as Anne Krueger points out, there are few models showing that link both rigorously and convincingly. Theorists have handled better the effects on trade of exogenously given growth than the effects on growth of different degrees of policy-determined openness. Real income levels can be shown to be modified by trade policies, but it is difficult to translate once-and-for-all trade gains into higher steady-state long-run growth rates. Little more than the poverty of extant growth theory should be read from all this.

As Krueger notes, there is substantial empirical evidence showing a positive correlation between export performance and growth. However, that evidence typically refers to post-World War II years; it could be argued that other historical epochs tell different stories. When most of the Third World was under colonial rule export promotion was the rule; growth consequences, especially when the industrialized countries were in a slump, were less than spectacular. Other development goals, such as those dealing with income distribution and domestic institution building, fared worse.

During the 1930's, Latin American countries which moved away from the gold standard and *laissez-faire* orthodoxies registered better growth rates than those which did not (or could not). Note that there is room for discussion whether the trade policies of the largest Latin American countries departed significantly from Krueger's criterion (equate *IMRT* to *DMRT*). The point is that when the industrialized world is in depression no country is small; external demand price elasticities sag, providing neoclassical justification for some trade taxes. Edgeworth, Keynes, and Krueger all may look approvingly upon Brazil's departure from *laissez-faire* during the 1930's, even if agreeing that policy could have been refined. One could conjecture, with *ex post* comfort, that even if one was aiming for "*IMRT* equals *DMRT*" perfection during the 1930's, it was less costly to err on the side of import substitution than of export promotion. An industrialized world riddled with unpredictable tariffs and import quotas could turn export drives into disasters.

Will the 1980's be more like the 1960's than like the 1930's? Some economists worry that fine tuning commercial policy to world circumstances will lead to renewed Third World excesses in import substitution. Perhaps, but the advice that Third World countries should design their trade policies as if the state of the world economy did not matter (or as if they were small at all times) suggests evangelical fervor rather than scientific analysis. The advice "equate *IMRT* to *DMRT*" is unimpeachable but, alas, not so operational once one departs from the small-country assumption and the number of goods multiplies.

Third World regional trading arrangements could be viewed as insurance against a rest-of-the-world in depression; the Latin American import substitution efforts of the 1930's would have been even more impressive within the framework of a regional common market. Krueger observes that under more prosperous circumstances southern common markets remain a complement to North-South trade, a complement whose economic welfare consequences are moot and whose politics are prickly.

Economic size is bound to influence the trade-growth nexus. Brazilian growth during this century has been vigorous under a variety of alternating trade regimes; Uruguayan or Dominican growth will show greater sensitivity to trade regimes and will be seriously impeded by excessive import substitution. What about social systems? Perhaps the most impressive bit of evidence in favor of Krueger's protrade arguments comes from the trend toward greater openness in socialist countries. Cuba, for example, views trade with both socialist and nonsocialist countries as a source of scarce foreign exchange needed to buy technology,

intermediate and capital goods, that is, as a key element in her growth strategy. In these matters Cuban planners go along with Ricardo and Chenery rather than with Samir Amin. Cuban growth has been modest, but seems superior to that of countries where a delinking approach has been seriously undertaken, as in Albania, Burma, and Kampuchea.

When all is said and done we remain unsure as to whether and when trade is the engine, the handmaiden, the brake, or the offspring of growth. Ingenious and mischievous economic historians can come up with examples of each. Conjectures that both trade and growth depend on third variables are attractive but difficult to test. But let me close on a positive note; if asked to devise trade policy for an *LDC* circa 1980, I suspect I would do something very similar to what Krueger would.

Regulation and the Choice of Prescription Drugs

By PETER TEMIN*

In 1938, shortly after the passage of the Federal Food, Drug, and Cosmetic Act of that year, the Food and Drug Administration (FDA) issued regulations creating "prescription drugs," that is, drugs that could be purchased only by prescription. There are many reasons to approve of such a regulation, some of which have developed since 1938, but I would like to consider some problems with it in this paper.

The problems caused by the prescription drug regulations are illustrated by a specific, tragic set of events which culminated in a lawsuit. Among the "wonder drugs" introduced in the 1940's was an antibiotic called chloramphenicol, sold exclusively by Parke, Davis as Chloromycetin under patent protection for the time considered here. At the time it was introduced, chloramphenicol seemed indistinguishable from the other broad-spectrum antibiotics, and it sold for the same list price throughout the 1950's (see U.S. Congress, 1961, Chart 10). But reports of fatal anemias associated with chloramphenicol surfaced in the early 1950's, and warnings to physicians were mandated by the FDA (see C. L. Lewis et al.; U.S. Congress, 1964, pp. 148–49). Chloramphenicol, it turns out, can cause two entirely separate kinds of anemias: a reversible anemia detectable while the drug is being administered and a fatal aplastic anemia that typically occurs after administration of the drug is terminated. There is no evidence that the incidence of aplastic anemia is related to the dosage of chloramphenicol, and relatively accurate estimates of the probability of contracting aplastic anemia after taking chloramphenicol—roughly one in 40,000—did not become available until 1969 (see Ralph Wallerstein et al.). The warning that the FDA insisted accompany the drug initially confused the two types of anemia, suggesting by its attention to blood studies undertaken during therapy that the fatal aplastic anemia could be so detected. Not until 1961 was the package insert changed to signal clearly the presence of two distinct risks, of which the more serious could not be anticipated by blood studies (see U.S. Congress, 1964, pp. 148–49).

Mary Ann Incollingo was born in late 1955. Her pediatrician prescribed chloramphenicol for respiratory infections on three separate occasions between mid-1958 and January 1960. Mary Ann's mother tried unsuccessfully to get an additional chloramphenicol prescription from the child's pediatrician in February 1960. Failing in this endeavor, she turned to her own physician, who prescribed chloramphenicol for the child without examining her on two separate occasions in early 1960. Mary Ann's mother gave her the drug as she thought Mary Ann needed it, ignoring the instructions on the prescriptions. Mary Ann contracted aplastic anemia some time before it was diagnosed in May 1960, and she died in 1962. Her parents sued their druggist for negligence; he joined the two doctors and Parke, Davis as additional defendants. Judgements against the two doctors and Parke, Davis were affirmed by the Pennsylvania Supreme Court in 1971 (see *Incollingo v. Ewing et al.*).

The Pennsylvania Supreme court's opinion in this case illuminates several of the

*Massachusetts Institute of Technology. This research was supported by a grant from the Sloan Foundation to M.I.T. on the Public Control of Economic Activity.

problems with the 1938 prescription-only regulation, both by where it found guilt and by the reasoning used to establish guilt.

First, there is the question of causation. In order to provide the occasion for damages, it was necessary to prove that the administration of chloramphenicol caused Mary Ann's death. The courts did not have to confront the question squarely, since "there was no real medical disagreement at trial that Chloromycetin (chloramphenicol) was in fact the underlying cause of Mary Ann's death." Not everyone who takes chloramphenicol dies. In fact, only a very few die. The mechanism by which the drug and the fatal anemia are connected is not known, and it therefore is not possible to trace an explicit causal chain from one to the other. Taking chloramphenicol changes the probability of contracting aplastic anemia. The best point estimate of the change is that the risk of dying from aplastic anemia rises from 1 in 525,000 to 1 in 40,000 (see Wallerstein et al.). The issue raised by the case therefore is whether changing a probability distribution qualifies as a "cause" of an event. Almost everything we do changes the probabilities that other things will happen, and this general definition of "causation" makes it hard to discriminate between possible "causes." The sense in which chloramphenicol causes aplastic anemia is very different from the sense in which an overdose of barbiturates causes death.

The pediatrician was judged liable in part because he failed to do the indicated blood tests. But while the blood tests are useful to identify one kind of anemia, they cannot anticipate the anemia that killed Mary Ann. The pediatrician consequently would not have learned anything about the eventual (proximate) cause of death from the test results. This was known by the time the case came to trial; but it does not appear to have been public knowledge at the time the drug was administered to Mary Ann. It was explained in the package inserts starting in early 1961, less than a year after Mary Ann's illness was diagnosed.

The argument against the pediatrician that he was negligent in not doing the indicated tests appears to rely on a model of drug action in which the dosage of the drug matters. But the dosage of chloramphenicol matters for the reversible anemia the blood tests detect; there is no evidence that the incidence of the aplastic anemia that killed Mary Ann is dose related. Alternatively, the argument might be that the doctor should have done the tests because knowledge at the time indicated them, even though subsequent knowledge showed them to be irrelevant. To state these arguments explicitly is to expose their weakness.

It should be noted that the package insert was directed at physicians. According to the FDA's 1938 regulation, information was not to be furnished to consumers. While this appears to have been a device to enforce the regulation, it had important implications for both producers and consumers.

Since prescription drug producers had a restricted market for their products, defining membership in the market by the ability to prescribe drugs, the drug firms could channel advertising directly to the people in that market. In addition to the usual forms of written and visual communication, the drug companies could and did send their personal representatives to talk to individual doctors. After all, doctors were both few in number relative to the total population and easily identifiable. The 1950's even may be characterized as the golden age of detail men, lying between the discovery of the new antibiotics and the 1962 Drug Amendments.

The case against Parke, Davis revolved around the firm's detail men. The printed warnings accompanying chloramphenicol to the doctor were deemed adequate, both because of their content and because they were in conformity with the FDA's directives. The question was whether the printed information had been nullified by the oral presentations of the detail men. The question therefore concerned a form of marketing fostered by the 1938 regulation.

Turning from the producer to the consumer, it is important that Mary Ann's mother was not found at fault, even though she had subverted the prescribing system in several ways. She had not acquiesced in the pediatrician's decision to stop giving her

daughter the drug. She had tried to get the pharmacist to sell her chloramphenicol without a prescription. She did not even follow the instructions on the prescription she finally obtained from her own physician. Despite the clear intent of the FDA's regulation to exclude her from the consumption decision, she made the operative decisions about whether and when to give her daughter chloramphenicol. Yet she was absolved of all responsibility for her actions by the existence of the 1938 FDA regulation.

The opinion in this case distorts reality (as we now know it) and thereby creates improper incentives. First, by ignoring the uncertain nature of chloramphenicol's action and by not taking adequate notice of the changing understanding of these actions, the opinion invites people to ignore the stochastic property of drug use. Second, by absolving the mother of all responsibility, the opinion and the regulation that underlies it encourages individuals to ignore unpleasant implications of their activities.

The regulation creating a separation between over-the-counter drugs (over which consumers have choice) and prescription drugs (about which consumers are supposed to know nothing) assumes that you can control drug decisions. It deals with drug risks by attempting to deny consumers both the choice of dangerous drugs and the knowledge upon which such a choice might be based. It tries to exclude individuals from this choice, rather than including them in it and educating them about it. The regulation therefore obscures consumers' responsibility and their need to cope with rare stochastic events.

Going further, the regulation created a class of ignorant consumers as well as a class of prescription drugs. By specifying that information on prescription drugs be expressed in ways intelligible only to doctors, the regulation tried to prevent consumers from avoiding the regulation. But, as the case under discussion shows, consumers cannot be excluded from the decision to buy drugs, much less from the decision to take drugs. The result of the regulation therefore is to make consumer's participation in these decisions ill-informed and, at least legally, irresponsible. Phrased differently, some part of the gap between the drug knowledge of the average doctor and the average consumer is the product of regulation. This needs to be kept in mind when the magnitude of the gap is exhibited as an argument for more regulation.

The context in which Mary Ann received chloramphenicol was very much a result of government policy as expressed through the FDA's 1938 regulation. The way in which drugs were marketed and the way in which a person obtained drugs by the late 1950's were products of this regulation as it interacted with the rapidly changing postwar drug industry. To understand this interaction, it is necessary to examine the history of the regulation.

It had been possible to utilize prescriptions before 1938, but it also had been possible to buy virtually any drug without a prescription. The 1938 Federal Food, Drug, and Cosmetic Act gave the FDA the right for the first time to deny a nonnarcotic drug access to the American market. The regulation created an intermediate category so that instead of choosing between making a new drug freely available or not available at all, the FDA introduced the additional possibility of making a drug available on a controlled, prescription basis. Second, the intent of the regulation, judging by early prosecutions under it, was to control the quantity of drugs that people took, under the presumption that drug dosage is important. And third, the FDA attempted to eliminate the consumer from the choice of prescription drugs by denying information about those drugs to him. This was in the first instance a way of enforcing the regulation—by denying consumers the means to circumvent it—and subsequently a means of transforming consumers into patients (see my 1979a paper).

The regulation creating prescription drugs was part of the general administrative regulations needed to implement the 1938 law. Its connection to the law, however, is tenuous. I have just given some reasons why the FDA might have wanted such a regulation, but these reasons do not constitute a legal base for the regulation. The fragility of the

regulation's legal foundation is suggested by a recent (October 1979) Federal District Court decision declaring the regulation invalid because it was not written in accord with the law, that is, because it did not implement the 1938 law as written (see *U.S. v. Colahan et al.*).

The drug industry changed dramatically in the years after 1938. World War II created a huge demand for anti-infective agents, which led to the preparation of therapeutically useful concentrations of penicillin and the discovery of manufacturing techniques for it toward the end of the war. Other antibiotics were discovered and marketed after the war, and research became an important factor in competition among drug companies. Drug firms increased in size and in the range of functions they performed. Many nonantibiotic drugs were introduced as well, and the modern drug industry—dominated by large firms and patented new drugs—began to emerge (see my 1979b paper).

The 1938 regulation had not specified which drugs were to be placed in the intermediate, prescription-only category, and the 1951 Durham-Humphrey Amendments that wrote it into law followed suit. The drug manufacturer could decide how to classify a drug, give the drug a label appropriate to its selection, and sell it, subject only to the FDA suing it for mislabelling its drug, that is, for providing the label for an inappropriate category. (This happened seldom, if ever.) The drug firms introduced virtually all of the new drugs as prescription drugs, apparently without sustained opposition. Since drug firms are engaged in profit-making activity, they undoubtedly thought they derived a commercial advantage from this designation. There are fewer doctors than consumers, and they can be reached easily through specialized information channels. In fact, the drug firms could send personal representatives, detail men, to each doctor to explain the merits of their products. In addition, as drugs became more powerful and therefore could be used to treat sicker people, selling through doctors became an attractive distribution path to consumers as well. And the choice also was consistent with the FDA's position of reserving difficult choices to professionals. Prescription drugs consequently increased rapidly as a proportion of drugs on the market, and the transitional class of drugs introduced in 1938 to mediate between the two major categories of drugs became itself the major category.

The regulation was issued in 1938 to resolve a particular legal quandary, to blur an excessively sharp—in the FDA's view—legislative distinction between marketable and nonmarketable drugs. At the time, it affected only a very few drugs and aroused almost no comment. But as the drug industry underwent a technological revolution after World War II, this marginal regulation moved to the center of the industry, affecting both the way drugs are produced and sold and the position that doctors occupy in their distribution.

It is tempting to imagine the mandatory use of prescriptions as the consequence of medical pressure on the regulatory process, and the result of the doctors ability to "capture" the FDA. But there is no indication that the process worked this way. No one anticipated the postwar use of the regulation in 1938. At that time, it looked like a small change having little impact, and no doctors even testified at the hearings on this regulation (see my 1979a paper, p. 100). The subsequent expansion of the regulation's scope was the result of changing drug technology and the drug companies' decisions on marketing strategy as approved by the FDA. Doctors were the passive recipients of this exclusive and valuable license to prescribe drugs.

It is not even possible to blame doctors for creating the obfuscations present in the way prescriptions are written. While doctors have been active participants in the professionalization of drug distribution, the initial impulse toward a sharp separation between professionals and laymen came from the FDA.

The FDA was not a party in *Incollingo v. Ewing et al.*, but it was an important influence in the case. Its regulation affected both the context in which Mary Ann was given chloramphenicol and the way in

which the court regarded her parents' suit. But, as I have tried to show here, the FDA's 1938 regulation looked quite different in 1959 or 1971 than it had in 1938. The number of drugs it affects, the organization of the drug industry, the role of doctors all had changed—partly to be sure, as a result of the regulation itself. To understand *Incollingo v. Ewing et al.*, therefore, it is necessary to trace the development of this regulation as it evolved from 1938 on.

This development suggests that the FDA may have gone too far, possibly inadvertently, in the direction of discouraging consumer involvement in the drug selection process. Providing consumers with more information about potent drugs, through measures like patient package inserts, would move back in the direction of promoting responsible consumer participation in drug choices. Educating consumers about the nature of risks could have more general benefits as well.

REFERENCES

C. L. Lewis et al., "Chloramphenicol (Chloromycetin) in Relation to Blood Dyscrasias with Observations on Other Drugs," *Antibiotics and Chemotherapy*, Dec. 1952, *2*, 601–09.

P. Temin, (1979a) "The Origin of Compulsory Drug Perscriptions," *J. Law Econ.*, Apr. 1979, *22*, 91–105.

______, (1979b) "Technology, Regulation, and Market Structure in the Modern Pharmaceutical Industry," *Bell J. Econ.*, Autumn 1979, *10*, 429–46.

R. O. Wallerstein et al., "Statewide Study of Chloramphenicol Therapy and Fatal Aplastic Anemia," *J. Amer. Medical Assn.*, June 16, 1969, *208*, 2045–50.

Incollingo v. Ewing et al., 444 Pa. 263 (1971).

U.S. Congress, Senate, *Report on Administered Prices: Drugs*, Senate rept. no. 448, 87th Cong., 1st sess., 1961.

______, **House,** *Hearings on Drug Safety Before the Subcommittee on Intergovernmental Relations of the Committee on Government Operations*, 88th Cong., 2d sess., 1964.

United States v. Colahan et al., U.S. District Court, Northern District of Ohio, Eastern Division, no. C78-1470A, "Memorandum Opinion and Order," Oct. 9, 1979.

The Market for Regulation: The ICC from 1887 to 1920

By THOMAS S. ULEN*

There is a great reassessment under way among economists concerning federal regulation of business. The first series of studies on regulation has almost invariably found that consumer welfare is less than it would be without that regulation. With the usual lag, these results have now begun to effect national policy to the extent that air and surface transportation have been substantially deregulated. A second series of studies is now appearing which goes beyond the measure of static costs and benefits of regulation to ask how we came to have the particular regulations that we do have. In this regard there are at least two contending theories (see George Stigler; Richard Posner). The economic theory of regulation hypothesizes that special interest groups, for whom the benefits of federal regulation are large, persuade Congress and the Executive to enact laws (for example, entry restrictions and minimum price schedules). Politicians are prepared to supply such regulation because its costs can be spread over so many voters as to make it almost unnoticeable while its benefits accrue to small, identifiable groups. Against this hypothesis is the public interest theory of regulation. The suggestion of this second hypothesis is that the government introduces regulation at the behest of consumers in order to maximize consumer welfare in industries where competition cannot be relied upon to do so.

I propose in what follows to study the origins and growth of federal railroad regulation to see to what extent these competing theories of regulation explain the historical record. It is also possible that neither hypothesis above can illuminate the area and that some third hypothesis will have to be proposed.

*University of Illinois.

I. The Demand for Regulation Before 1887

When President Cleveland signed the Interstate Commerce Act in February 1887, the era of *laissez-faire*, which had flourished only intermittently in the nineteenth century, ended. Thereafter, the issue of whether government should regulate independent business was dropped. Instead the question became how to make regulation more effective and, after 1920, how to extend it to other industries. As work by Oscar Handlin, Louis Hartz, Jonathan Hughes, and Morton Horwitz has demonstrated, this more pragmatic approach was indeed the time-honored one: government and the public never doubted its determination to regulate, but did differ at times on who should be regulated and how it should be done.

The rapid extension of the railroad network into new regions where settlement was sparse, and the increasing density of the rail service in older, more heavily populated regions in the decade after the conclusion of the Civil War has been much remarked. These developments in surface transportation in the early 1870's caused unease among three separate groups. These same groups constituted the principal demanders of federal railroad regulation:

1) The farmers in the upper Midwest and on the edges of the Great Plains were persuaded that they faced, in the form of the railroad, a monopoly. In this case farmers felt that through rates on interstate shipment of grain were excessive, thus lowering farm incomes below what they would have been had there been more competition. They demanded legislation from their state governments limiting the rates which could be charged by railroads and by warehouses, which were often owned by railroads. Although upheld in the famous *Munn v. Illinois* decision, this attempt to regulate in-

terstate commerce had already been abandoned in all the Granger states save one by 1877. The reason for this was that railroad investment shunned those states with the most restrictive legislation, and it became clear that monopolistic railroads were better than no railroads at all.

2) A second group which grew restless with the railroads in the 1870's was merchants and farmers in the East. As Lee Benson has noted, merchants in New York City felt that price discrimination by railroads on freight hauled into and out of the nation's largest port harmed their business as transhippers. Their allies in opposing the roads were farmers in upstate New York, who also felt themselves to be victims of discriminatory rate making in that their ton-mile rate was greater than the rate paid by farmers further west. Together these groups persuaded the New York State Assembly to hold the important Hepburn Committee hearings in 1879, whose results influenced Congress' actions in the mid-1880's.

3) The last group which became exercised as the rail network became more dense was the railroads themselves. More railroads in a given area meant more competition and thus lower profits. At first informally, the railroads tried to collude on prices charged, but these loose conspiracies of the early and mid-1870's fell prey to secret rate cutting by the members, particularly so during the depression after the Panic of 1873. From the beginning of the recovery in late 1877 railroads began to develop more sophisticated collusions. This development served, when effective, to increase the anger of the consumers of rail services.

The interests of these three groups were different. Therefore, when they expressed a demand for legislative relief, the sorts of legislation that they favored and that might have resulted ought to have been different. In principle there are three different bills being demanded. This is strictly true if only one of the group is to succeed. To the extent that none succeeds fully but each groups succeeds a little bit, the resulting legislation may well be internally inconsistent. The farmers in the West would have had as their main goal the lowering of rates on through freight. Eastern merchants and farmers would have been most concerned to have short-haul rates fall through legislation outlawing price discrimination. The railroads would have had two different strategies in this market of regulation: if they felt confident of being able to outbid and exclude the other interest groups, then they would have wanted legislation which reduced the costs to them of earning near-monopoly profits through collusion. They might have desired, for instance, that the government act as a manager for the collusion or that cheating on the cartel contract be punishable as simple breach of contract. If, however, the railroads had strong doubts about their ability to exclude other special interests, they would have chosen a more defensive strategy. In that case, the best that the roads might have been able to achieve would have been a watering down of the restrictions which the other groups might have succeeded in achieving.

The interactions among these three groups in the years just before and after the passage of the Interstate Commerce Act in 1887 reveals much about the origins and growth of federal regulation of independent business. The story of the farmers in the Midwest has been admirably told by George Miller and I. L. Sharfman; that of the Eastern merchants and farmers by Benson. I shall confine my attention in what follows to the activities of the railroads. Additionally, I shall look in some detail at the legislation which Congress supplied in 1887 to see if the provisions of the Act allow one to draw inferences about which interest group or groups succeeded in the bidding before Congress. Lastly, the amendments to the Interstate Commerce Act made in the early years of the twentieth century and in the Transportation Act of 1920 will suggest a third hypothesis about the origins and growth of regulation.

II. The Railroad Interest in Federal Regulation

In the *First Annual Report on the Internal Commerce of the United States 1876*, Joseph

Nimmo remarked that railroad cartels—called "pools"—were operating between all major regions of the country and had been doing so for ten years. Among the cartels upon which Nimmo reported, two stand out. The Southern Railway and Steamship Association had been established in 1876 by railroads and steamers in the South to control rates and market shares for the shipment of cotton from the interior to all ports in the South. An innovative technique adopted by that cartel in order to compel adherence to its rates and market-share allotments was the joint ownership by all members of the rolling stock used in through shipments of freight. The second late-nineteenth century railroad cartel of note is the Joint Executive Committee (*JEC*). That collusion embraced the major trunk lines and their many independent feeders between St. Louis and Chicago in the West and the eastern seaboard ports north of Baltimore. It is this second cartel upon which I wish to concentrate.

The *JEC* was the successor to a series of informal rate-setting agreements among freight agents of the trunk lines. As the received theory of cartels would predict, these informal arrangements were not successful, being repeatedly broken by secret cheating. The organization which appeared in June 1879 was a different matter. The colluders had experience with unsuccessful cartels, knew the shortcomings, and deliberately set out to erect internal enforcement mechanisms which would bring them higher profits with more certainty. The cartel took the following steps: 1) it compiled and published statistics weekly on the amount of flour, grain, and provisions carried by each member; the figures were verified by independent agents of the Chicago Board of Trade and only once did a member complain of their inaccuracy. 2) The *JEC* retained a Board of Arbitrators to hear testimony about and decide internal cartel disputes. The arbitrators were paid a retainer of $10 thousand per year and included such famous men as Thomas Cooley, Charles Francis Adams, and David Wright. 3) From March 1881, the cartel granted the commissioner, Albert Fink, the power to match immediately any price cut by a member. The effect of this device was to allow the loyal members of the collusion to act as a bloc—thus preserving some structure in the cartel—and to reduce the attractiveness of unilateral cheating. 4) The cartel empowered the commissioner to impose various economic sanctions on a discovered cheater. All or part of good faith deposits were forfeited upon a proven allegation of cheating, and for the smaller roads which cheated, the loyal cartelists refused to transfer freight with them. 5) Lastly, the cartel, at regular intervals, diverted freight among its members such that roads which were above their market-share allotments sent freight to those roads which had carried a share of the cartelized output less than that specified in the agreement.

The effect on cartel stability of these elaborate internal enforcement devices was remarkable. In the seven years before the passage of the Interstate Commerce Act in 1887, the cartel was successful in maintaining near-monopoly rates in more than three-quarters of the 328 weeks surveyed. Cheating and a breakdown of the cartel agreement occurred in fewer than one-quarter of the weeks surveyed (see my dissertation). Moreover, cartel success fluctuated with the demand for the transport services of the cartel. Cheating occurred when business was tapering off, usually because of the entry of new firms or because of a modest harvest in the Midwest and the Plains. Success in jointly monopolizing profits was most likely when demand was expanding.

In light of these findings, it seems less likely that the railroads had an exceptional interest in aggressively securing the government's aid in managing their collusion. Although it is possible that the monopoly profits foregone during the relatively infrequent periods of cheating were substantial, it is more likely that they were not substantial enough to warrant a concerted effort to bid for congressional power. This is especially so when one recognizes that there is little a priori reason to believe that the resulting legislation, even if ideally framed, would have necessarily been more effective

than the internal enforcement devices used before 1887 to compel adherence to the cartel contract. The punishment meted out for disobeyance of a statute law would not necessarily have been more of a deterrent to cheating than was the sort of punishment which the cartel was already meting out.

That an Act to Regulate Commerce was not necessary to the relatively stable operation of railroad cartels is borne out further by examining the provisions of the Act. Clauses 2 through 7 of the Act dealt with the sort of cartel agreement that the railroads had worked out formerly among themselves. None of these said anything specific about the legality of collusion, that being thought better handled by the common law, or, later, by the Sherman Act. One device which the *JEC* had relied upon for the smooth functioning of its collusion—the transfer of tonnage to equalize actual and allotted market shares—was specifically out-lawed in section 5. It should not be thought, however, that this illegality of pooling freight or revenue caused more instability in the cartel. Instead, a new means of promoting loyalty was developed. A road which was carrying less than its contractually allotted market share was allowed, by the other members of its rate bureau, to announce a lower rate until the resulting increase in business to that road equalized its actual and allotted shares. None of the cartels' other internal enforcement devices was affected by the Act. Nor can any of the other provisions in sections 2 through 7 be said to have greatly affected the probability of cartel stability.

And yet, in the years just after the passage of the Interstate Commerce Act the success of the cartel increased. Since the above considerations suggest that the presence of the Interstate Commerce Commission was not the factor responsible for this increased success, an alternative explanation must be found. The one which best fits the facts is this: the ICC was, like many first efforts, ineffectual; it neither greatly increased railroad profits nor consumer's surplus. Instead, the same factors which had led to stability in railroad collusion before 1887 accounted for it after 1887. The shipment of grain, flour, and provisions from the Midwest to the East boomed between 1887 and 1893. Just as importantly, entry into railroading, as measured by the volume of investment in the industry, fell continually and rapidly from 1886 to the end of the century. When the demand for transport services was sharply curtailed by the severe downturn of 1893–96, the railroad collusions broke down in much the same fashion as they had when demand fell before 1887.

III. The Growth of Regulation After 1896

During the depression that lasted from 1893 to 1896, approximately one-fifth of all the rail mileage in the country went into receivership. Even among the healthy roads, the pressures of decreasing demand were so severe that the resulting rate wars were said by contemporaries to have been the worst of the late-nineteenth century. This set of circumstances, when combined with an ineffectual federal regulatory agency and a Justice Department which already had filed suit against several railroad cartels for violation of the Sherman Act, led the railroad industry to seek alternatives to collusion. The merger of independent businesses was the means adopted. In the trunk line territory between Chicago and the Atlantic ports, the eight independent roads of the early 1890's combined into three, with strong financial interests overlapping in two of those.

Almost as if to put its seal to the end of an era, the Supreme Court held in *United States v. Trans-Missouri Freight Association* and in *United States v. Joint Traffic Association*, that railroad collusion of the sort practiced since 1887 violated section 1 of the Sherman Act. These decisions meant that there would be no turning back for the railroad industry. They had opted for the "merger to monopoly" route, creating altogether different regulatory problems than those which had agitated the public in the 1880's.

This chain of events must be traced in order to understand the subsequent efforts by the Congress to breathe life into the ICC. In 1903 the Elkins Act strengthened the

penalties for discrimination and for nonposting of rates. Three years later the Hepburn Act gave the commission the power to set maximum rates, made thirty-day prior notice of rate changes the law, and imposed treble damages on those found guilty of receiving rebates. The Mann-Elkins Act of 1910 cleared some of the terrible confusion which had attended the 1887 Act's long haul, short haul clause and also empowered the ICC to suspend proposed rate changes. In these acts it is difficult to discover the interests of any single group being represented, other than possibly the ICC itself. As Albro Martin has shown, the more efficient railroads were not served by these provisions. Nor has the case yet been made that the public interest was much served by these extensions of regulation. That something else was precipitating refinements of the regulations dealing with surface transportation must be more closely considered. This seems the more pressing in view of the capstone of this edifice of regulatory legislation, the Transportation Act of 1920. Congress gave the ICC the power to specify minimum rates, to determine entry and exit from the railroad industry, to specify the amount and kind of capital formation which roads might undertake, and to preside over matters of consolidation. In a gesture to the original Act to Regulate Commerce, the ICC was granted the power to approve the pooling of freight and revenues.

IV. Conclusion

The history of federal railroad regulation between 1887 and 1920 cannot be fully comprehended using either the public interest theory or the economic theory of regulation. The Act to Regulate Commerce cannot be said to have resulted in a consistent piece of legislation. Bidding for congressional action is not a zero-sum game in which, in this instance, only one of the three groups identifiable as interested in action will have its desires alone written into law. Each group got something in the Act, which made the result palatable in the short run. But as circumstances in the economy changed over the long run, the inconsistencies of the original regulation became evident and adjustments were made. It is not at all clear whether any of the original interest groups continued to express a demand for these alterations. Indeed, there is evidence that none of these had a clear interest in extending and tightening the regulation to the extent that resulted in 1920. The only group which may have had sustained interest in the events from 1903 on was the ICC itself.

REFERENCES

Lee Benson, *Merchants, Farmers, and Railroads*, Cambridge, Mass. 1955.

Oscar Handlin, *Commonwealth: A Study of the Role of Government in the American Economy, Massachusetts, 1774–1861*, New York 1947.

Louis Hartz, *Economic Policy and Democratic Thought: Pennsylvania, 1776–1860*, Cambridge, Mass. 1948.

Morton J. Horwitz, *The Transformation of American Law, 1780–1860*, Cambridge, Mass. 1977.

Jonathan R. T. Hughes, *The Governmental Habit*, New York 1977.

Albro Martin, *Enterprise Denied*, New York 1971.

George Miller, *Railroads and the Granger Laws*, Madison 1971.

R. A. Posner, "Theories of Economic Regulation," *Bell J. Econ.*, Autumn 1974, 5, 335–58.

I. L. Sharfman, *The Interstate Commerce Commission*, New York 1931.

G. J. Stigler, "The Theory of Economic Regulation," *Bell J. Econ.*, Spring 1971, 2, 3–21.

T. S. Ulen, "Cartels and Regulation," unpublished doctoral dissertation, Stanford Univ. 1978.

Munn v. Illinois, 94 U.S. 113 (1877).

United States v. Joint Traffic Association, 171 U.S. 505 (1898).

United States v. Trans-Missouri Freight Association, 166 U.S. 290 (1897).

U.S. Bureau of Statistics, *First Annual Report on the Internal Commerce of the United States 1876*, "Part Second...on the Commerce and Navigation...," Washington 1877.

Regulation, Deregulation, and Economic Efficiency: The Case of the CAB

By JOHN C. PANZAR*

The current wave of deregulatory enthusiasm makes it easy to forget that the creation of the CAB and its regulatory authority was itself a result of the political and economic forces which prevailed a relatively short time ago. Since deregulation nowadays is viewed as a move toward economic efficiency, it is tempting to argue that it represents a triumph of the public interest over the special interests which, presumably, were responsible for the formation of the CAB cartel. This would be a serious oversimplification, because 1) it is possible to view the *CAA* of 1938 as a response to a potential market failure *then*; 2) deregulation may give up the possibility of achieving optimal, first best resource allocation *now*; and 3) regulatory machinery which could have been used to pursue either economic efficiency or cartel profit maximization clearly did neither. While the last point requires no elaboration, the first two are somewhat novel, and rather provocative.

The concept upon which most of my arguments are based is that airline markets are, even today, more accurately described as being structurally contestable rather than competitive. Even the strongest proponents of deregulation concede that few air traffic markets can efficiently support more than a small number of firms actually operating *in* the market. (Forty-five years ago this number was almost certainly one.) Efficient market performance is expected to result from the large number of *potential* entrants and the easy entry and exit dictated by the technology, in other words, because the market is easily *contested.*[1]

I. The Airline Industry of the 1930's and the Civil Aeronautics Act of 1938

Let us attempt to understand and evaluate the *CAA* of 1938 from the perspective of an economic analyst in 1935, when the report of the Federal Aviation Commission (*FAC*) was transmitted to Congress and the drafting of the Act began. Assume that the analyst is familiar with modern regulatory theory and the development of the industry up to that time, but blissfully ignorant of our recent experience of forty years of CAB regulation.[2] Given the premise that, for most routes, it was economic for only one carrier to transport the mail and impossible for an airline to be viable carrying passenger traffic alone, airline markets can be considered to have been joint-product natural monopolies. This posed potential problems for the design of an institutional structure capable of fostering economically efficient development in the industry.

Judging from the legislative history, Congress viewed various manifestations of "excessive competition" as the most serious potential problem. The most striking expressions of this concern are the ringing denounciations of competitive bidding for

*Bell Laboratories. The views expressed are my own and do not necessarily reflect those of Bell Laboratories or the Bell System. I would like to thank Robert D. Willig for helpful comments and suggestions and Elizabeth Bailey and Michael E. Levine for pointing me in the right direction. They are not responsible for where I ended up.

[1] William Baumol, Robert Willig and I study this concept in detail in "The Theory of Value in Contestable Markets," (monograph in preparation).

[2] An invaluable source of quantitative information on the industry *before* regulation is contained in Robert Serling's compendium of airline schedules, 1929–39. An excellent guide to the legislative history of the Act, complete with excerpts from Congressional hearings, is provided in *Civil Aeronautics Board Practices and Procedures* (hereafter, the Subcommittee report).

airmail contracts in the congressional debate. At first, this seems somewhat surprising. While economies of scale might dictate that there be only one firm operating *in* the market, competition *for* the market might have been effective.

Michael Levine (1975) has argued that Congress and the public were overly influenced by the disastrous results of earlier competitive bidding regimes. Prior to 1934, the high-handed behavior of Postmaster General Brown as well as the wide latitude permitted him by the Watres Act, made that period one of *de jure* rather than of *de facto* competitive bidding. After repeated scandals, President Roosevelt cancelled all existing mail contracts and ordered the Army to carry the mail. That disastrous experiment led to the establishment of the *FAC* and to yet another attempt at letting airmail contracts by competitive bidding. That temporary measure was transformed into a curious regulatory regime in which the U.S. Post Office awarded the contract to a low bidder, but the ICC was mandated to adjust rates to fair and reasonable levels. Ludicrously low rates ($.001/ton mile) were bid, since the airlines fully expected that the mail rate would be raised, and made due in the interim by exploiting the (effective) passenger service monopoly conferred by the airmail contract. Thus Levine concluded that "...Congressional attempts to use competitive contracting were doomed to failure by lawlessness and incompetent institutional design..." (1975, p. 323). In other words, the concept had not been given a true trial.

There was, nevertheless, a valid reason why routes could not have been *efficiently* allocated by competitive bidding. Passenger transport, while still not generally viable on a stand-alone basis, was already emerging as the more important of the joint products provided by the industry. A competitive bidding system for letting airmail contracts, even if it worked perfectly, would succeed only in eliciting the mail rate which would yield zero profits in conjunction with *monopoly* pricing of passenger service. This will never lead to socially optimal prices if there is any elasticity of demand for air travel. This unfortunate situation would persist until either passenger service became viable on its own or it became economic to split the mail among two or more carriers; that is, until markets were no longer natural monopolies. In this environment it was sound economic policy to consolidate the responsibility for determining mail rates and passenger fares in a single entity.

While the above argument, combined with some minimal concern over monopoly exploitation, can be used to explain the rate authority given to the CAB, one must still find a rationale for the certificate of public convenience and necessity which the Act required to enter an airline market. It is clear that the industry wanted this provision to protect their markets from "fly-by-night companies" and "chiselers"; that is, those willing to provide service at a lower rate. (See the testimony of Colonel Edgar S. Gorrell, president of the Air Transport Association, as quoted by the Subcommittee report pp. 210–12.) But what reason did Congress have for incorporating this provision into the statute?

The reasoning of the highly influential *FAC* report is quite clear on this point (see the Subcommittee report, p. 213). Certificates of convenience and necessity were to be required for two equally important reasons: to ensure safety and to prevent excessive competition. The authors' concerns on the latter score seem to spring from the common carrier functions that might be required of certificated carriers. They refer to the possibility of cream skimming by fly-by-*day* and seasonal entrants which would be able to forgo the additional expenditures required to fly at night and/or maintain year-round service.

Recent investigators do not seem to have taken this explanation seriously. The Subcommittee report, for example, points to "...other vaguer concerns such as a presumed similarity between railroads and airlines..." (p. 207) as being the driving forces behind the *CAA*. While this conclusion may be partially due to the contradictory impression created by the *FAC* report's equally clear *pro*competitive language, I be-

lieve historical factors also play a role: in this instance the prevailing opinions of regulatory economists at the time the Subcommittee report was written in 1975.

After the Great Depression, the issue of cream skimming was minimized to the point where it became almost "axiomatically" impossible. Alfred Kahn succinctly summed up this conventional wisdom: "If a natural monopolist is producing and pricing as efficiently as possible, there is no need to bar competitive entry: it is economically unnecessary and will not take place anyhow" (p. 223). However, recent research on pricing and entry in regulated markets has made it clear that there may exist situations in which it is *impossible* for a firm (or regulator) to set prices which do *not* provide profitable opportunities for cream skimming by potential entrants, even when the firm is operating under natural monopoly conditions and producing and pricing efficiently. The natural monopoly may be *unsustainable.*

To illustrate, consider a single product market in which the demand curve intersects the average cost curve within the natural monopoly region but beyond its minimum point. Any viable price quoted by a monopolist required to serve all demand can always be undercut by a cream skimmer operating at minimum average cost. (See my paper with Robert Willig for a discussion of this and more complicated, multiproduct examples of unsustainability.) The end result in such markets is problematic; what is clear is that *no free entry noncooperative (Nash) equilibrium exists.* Depending upon the dynamics built into the model, an endless cycle of price cutting, entry, and exit could theoretically result. Such markets readily might be described as disorderly and chaotic. (With a single product one way out of this dilemma is contract bidding. The firm awarded the contract would be *protected* from entry and *required* to serve all demand at the quoted price. However, as argued earlier, this simple approach is incapable of efficiently dealing with the case of multiple products.) Ironically, the contestability of airline markets, which makes monopoly exploitation difficult, makes it more likely that unsustainability will be a problem.

This argument, in conjunction with equally clear recommendations that desirable entry *not* be impeded, suggest that Congress was merely being prudent by requiring certificates of public convenience and necessity and that it intended to rely upon the judgment of an expert agency to determine a balanced entry policy. Senator Truman's confession (quoted on p. 81 of the Subcommittee report) that it was, unfortunately, necessary to *trust* the Board not to block the entry of new firms into the industry seemed incredibly naive just a few years ago. But that was before Kahn, Bailey, and Levine arrived upon the scene.

II. Regulatory Reform and Deregulation

Recent history would seem to indicate a clear triumph of economic efficiency over the "dead hand" of regulation. While I would be among the last to suggest that deregulation was a mistake, I argue that the issues are more complicated than is commonly thought; since, if anything, recent results in regulatory theory suggest that regulation by enlightened, but not omniscient, regulators could in principle achieve greater efficiency than deregulation. Thus, this solution also involved a tradeoff between pragmatism and theoretical optimality.

Academic criticism of the CAB tended to be of two types, typified by Levine (1965) and by George Douglas and James Miller. The first group questioned the need for *any* governmental *economic* regulation of the airlines, by appealing to the impressive performance of the (relatively) unregulated California intrastate market. In contrast, the Douglas-Miller approach was evolutionary rather than revolutionary. They proceeded by carefully building analytical theoretical models of industry behavior under the CAB's price and entry regulation and deriving recommendations for a more enlightened regulatory policy. A result of this division of labor was that very little attention was given to a *theoretical* analysis of unregulated airline markets.

The problem is interesting and nontrivial because of the monopolistically competitive nature of airline markets. Product differentiation is an unavoidable result of the variation in flight departure times, and the effects of flight frequency and load factor on service quality.[3] The basic results of my analysis are that 1) when the direct benefits (to consumers) of increasing flight frequency are exhausted, socially optimal choices of price and frequency result in zero profits for the industry; but 2) a noncooperative, free entry equilibrium always results in higher prices, lower load factors, and greater frequency than are socially optimal, and the employment of inefficiently small aircraft.

However, much of this inefficiency can be overcome by regulating *maximum* price and allowing free entry. Furthermore, these efficiency gains do not require an omniscient regulator, since in many interesting cases the optimal price is that which, with free entry, results in the maximization of passengers carried. Thus the optimum could, in principle, be achieved iteratively using readily available data.[4]

Why then was deregulation chosen over regulatory reform, and *how* was it achieved? Deferring the first issue to the conclusion, let me briefly summarize the "recipe" for successful deregulation put forward by Elizabeth Bailey. Academic criticisms of CAB regulation were reinforced by the thorough analysis of the Subcommittee report, which marshalled persuasive evidence to counter the objections and dire predictions put forward by the Board and the industry. Some protection for special interests who figured to suffer from deregulation, such as labor and small communities, was incorporated into the legislation. Finally, under Kahn, the board initiated a strongly procompetitive policy which served to convince Congress of the feasibility of competition and to prompt some carriers to feel that, shorn of the Board's *protection*, they would be better off without its *restrictions*.

From this argument and the upsurge in air travel and industry profits, it is tempting to conclude that (nearly) everyone has been made better off and to speculate that such a Pareto improvement is almost a necessary condition for changing the status quo. Let me inject a note of caution. *Basic* coach fares have risen steadily under deregulation. The gains in consumer's surplus and profits achieved by the proliferation of (discriminatory) discount fares are partially offset by the costs born by business travelers due to higher load factors. They constitute a large group of travelers harmed by deregulation without compensation. (The extent of their losses due to stochastic delay can in principle be quantified using the methodology developed by Douglas and Miller.)

III. Conclusions

What can be learned from this brief survey of regulatory entry and exit? More specifically, what accounts for the contrast between public policy toward (inherently) imperfect airline markets in the 1930's with that in the 1970's? I submit that the differences resulted from the interplay of three basic forces. 1) A change in public attitude. Competition was viewed with as much suspicion then as regulation is today. This shift is, in part, due to the superior relative performance of competitive airline markets in the interim. 2) The increase, due to market growth, of the efficient number of firms *in* most markets from one to two or more. 3) The basically contestable nature of airline markets, which implies that, if anything, entry will tend to be excessive rather than insufficient. When most markets are natural monopolies, entry *might* lead to *serious* difficulties with regard to nonexistence of equilibrium and subsequent loss of service. When natural monopolies are not present,

[3]Douglas and Miller (p. 178) recognized these issues and cited an unpublished paper by Douglas yielding results consistent with those reported here, which are from my 1979a paper.

[4]This simple, myopic regulatory rule also emerged from the analysis of the somewhat different models employed by Arthur DeVany and by me in my book.

entry will tend to cause only the (presumably) minor inefficiencies usually associated with monopolistic competition. Thus not only has the regulatory process failed to live up to its *theoretical* potential, the "downside risks" of the competitive alternative have been sharply reduced.

REFERENCES

E. Bailey, "Deregulation and Regulatory Reform: U.S. Domestic and International Air Transportation Policy During 1977–79," in Paul Kleindorfer and Bridger Mitchell, eds, *Regulated Industries and Public Enterprises: European and United States Perspectives*, Lexington 1980.

A. DeVany, "The Effect of Price and Entry Regulation on Airline Output, Capacity, and Efficiency," *Bell J. Econ.*, Spring 1975, *6*, 327–45.

George Douglas and James Miller, *Economic Regulation of Domestic Airtransport: Theory and Policy*, Washington 1974.

Alfred Kahn, *The Economics of Regulation: Principles and Institutions*, Vol. II, New York 1970.

M. Levine, "Is Regulation Necessary? California Air Transportation and National Regulatory Policy," *Yale Law J.*, July 1965, *74*, 1416–47.

______, "Regulating Airmail Transportation," *J. Law Econ.*, Oct. 1975, *18*, 317–59.

John Panzar, (1979a) "Equilibrium and Welfare in Unregulated Airline Markets," *Amer. Econ. Rev. Proc.*, May 1979, *69*, 92–95.

______, (1979b) *Regulation Service Quality and Market Performance: A Model of Airline Rivalry*, New York 1979.

______ **and R. Willig,** "Free Entry and the Sustainability of Natural Monopoly," *Bell J. Econ.*, Spring 1977, *8*, 1–22.

Robert Serling, *Birth of an Industry*, Chicago 1969.

U.S. Congress, Senate, Committee on the Judiciary, Subcommittee on Administrative Practice and Procedure, *Civil Aeronautics Board Practices and Procedures*, Washington 1975.

DISCUSSION

ROGER G. NOLL, California Institute of Technology: Peter Temin's work makes the important point that following the 1938 Food, Drug, and Cosmetic Act, consumers have had to use doctors as gatekeepers for access to most drugs. This distorts incentives through the medical care sector, and creates numerous potentially important inefficiencies. Temin has added two new observations: first, the post-1938 regulations make consumers more ignorant about drugs than they would otherwise be, creating evidence in support of further regulation; and second, the regulations have adverse effects upon the attitudes of courts in adjudicating damage suits. A third point made by Temin is that the 1938 changes probably made the drug industry less competitive; however, this is a secondary point and is (correctly) not claimed as original.

Most economists would not quarrel with Temin's policy conclusion that government should rely more upon consumers in the process by which drug decisions are made. Indeed, the FDA in recent years has advocated better information for consumers about prices of drugs and their aspects. A gap is left in Temin's analysis by the conclusion to eliminate mandatory prescriptions. His proposal of patient package inserts is normally suggested in the context of the traditional mode of medical practice, not as a substitute for prescriptions. It is one thing to say that patients know their own attitudes about alternative risks better than anyone else, including doctors, but it is another to say that *PPI*s and the like will enable patients to make unassisted decisions about drug choice and dosage.

Temin correctly points out that in 1938 the problem of drug choice was not very complicated for two reasons: there were not very many useful drugs, and doctors did not know very much about the effects of the ones that were then in use. In this milieu, one could hardly argue that the disenfranchisement of drug consumers accounted for some quantum leap in the welfare of the populace. Temin has argued elsewhere that the 1938 changes in FDA regulation were simply part of a general trend of the times to seek to interpose government between participants in a market owing to the decline in faith in markets that came about as a result of the Great Depression.

Temin notes that a more complicated world developed in the 1950's, which naturally raises the question of whether the mandatory prescription regulation is justified now, even if it was not in 1938. For some categories of drugs, Temin is undoubtedly correct. For example, one could imagine a world in which a patient could deal directly with a medical laboratory to obtain tests about the cause of an infection, and if the laboratory identified a particular bacterium as the cause, then deal directly with a pharmacist to purchase an appropriate antibiotic. A visit with a doctor may not always be a necessary part of the process. However, doctors are aware of some types of problems and uses of drugs, and particularly drug interactions, that consumers might miss, even if we all owned our own *Physicians' Desk Reference*, as is emphasized by Temin's horror story about chloramphenicol.

Even if this were not the case, economists' views about the reliability of informed choice are not widely shared in the case of medical care. Doctors and public health officials note that a very large proportion of drug use is by either children or the elderly, neither of which is necessarily able to make informed judgments. Moreover, medical care professionals claim that alarmingly large numbers of people abuse drugs or suffer from hypochondria, and argue that doctors keep some of them away from self-destruction by use of the prescription regulation. These views cannot simply be ignored in the debate about the proper scope of drug regulation.

Temin's discussion of the court case involving chloramphenicol leaves a few questions open. The efficiency consequences of the case that Temin discusses seem to not be very important. Consider an automatic rule

that caused drug companies to pay exact compensation (if it could be calculated) for all deaths of patients taking a particular drug and dying from a particular cause that was known to be much more likely among those taking the drug. This would be as efficient as a strict *caveat emptor*, except for problems resulting from interpersonal differences in risk preferences. In essence, this situation is one in which medical damage insurance is bundled with the potentially damaging product. From the point of view of the firm, correct compensation internalized the secondary effects of the drug into its cost structure. The fact that the event is probabilistic does not necessarily create an inefficiency unless the incidence of the disease from other causes is large relative to the expected increment in damages due to the drug. In the case cited by Temin, the likelihood that the drug in question caused the disease that killed the patient was an order of magnitude greater than the likelihood of all other causes combined.

The efficient liability rule, of course, would be to assign damages equal to the product of total damages and the probability that a particular drug caused the death conditional on the death occurring. Of course, the legal system is probably incapable of developing such a rule, and in instances in which there are multiple possible causes of similar likelihood, the inefficiencies of total compensation can loom large. This is, indeed, a good argument against use of the civil liability system to deal with this kind of problem.

Finally, Temin's observations on the problems with the chloramphenicol case are attributed to the 1938 changes in regulation that absolved consumers from any responsibility to take care of themselves. This seems to me to be implausible, for similar cases abound in all facets of civil liability case law, not just drug-related cases.

Temin is right in noting that in the chloramphenicol case a rather silly posture was adopted regarding probabilistic events and the responsibility of a consumer. But this posture strikes me as too common a feature of civil law to be regarded as causally connected to the particulars of drug regulation. Like the 1938 changes themselves, this decision was a small part of a much larger movement in the law.

Ellis W. Hawley, University of Iowa: Thomas Ulen's discussion of the "market" for railroad regulation pulls together much of what has been established about the origins and early history of the Interstate Commerce Commission. It suggests that the ICC itself may well have been the only beneficiary of this regulatory growth and that possibly it was one of the principal forces behind it. Here it is offering hypotheses for testing rather than conclusions based on hard data. But the suggestions are intriguing and seem in line with recent tendencies toward seeing regulatory operations as "businesses" or "industries" working actively to broaden and sustain the demand for their products.

The principal weakness of Ulen's paper in my judgment is its seeming ignorance of a substantial body of historical scholarship illuminating the workings of railroad politics and the forces behind regulatory change in the years 1903–20. K. Austin Kerr, in particular, has detailed the continuing initiatives of particular interest groups, showing how the legislation and regulatory action of the period continued to be shaped by competing regulatory demands, how external conditions favored first one interest and then another, and how the rise of the railway brotherhoods injected another major force into the battles between carriers, merchants, and farmers. Also entering the picture were groups concerned with political and social evils, and convinced that proper regulation of railroad behavior could help to eradicate them.

One major concern of this sort, for example, was with power that had arisen and established itself outside the political system. If democratic institutions were to survive, it was believed, such power had to be humbled and rendered accountable to political authorities. And if this was costly, the costs were justified by the political

benefits to be gained. There was a market here for regulation as a political good if not an economic good. Another major concern was with the growth of social disorder and class conflict, with what was perceived as the loss of an earlier harmony and community spirit. It was to this concern that the efficiency movements of the period, with their visions of a restored harmony through scientific management and engineered abundance, were addressed. For them, regulatory reform became a way of allowing the new harmonizers to engage in social reconstruction, and for them regulation was a social good.

Initially, this scientific management idea was urged upon the railroads as a way whereby they could meet rising costs and secure new revenues without increases in railroad rates. This was Louis Brandeis's argument in 1910 and one that the ICC reiterated in its rulings in 1914. From 1916 on, however, the railroads were able to turn the idea to their own advantage, arguing that only through fundamental reorganization of the regulatory system, meaning substantial scrapping of the ICC controls that they disliked, could the gains promised by scientific management be secured. This was an argument that helped the railroads to win powerful allies in the Wilson Administration and even to win over a segment of the shipper community. It was an argument that gained further credence when unified federal operation became necessary to meet war demands. And it was an argument that retained enough power to shape important sections of the Transportation Act of 1920. Indeed, this act is best seen not as a further augmentation and tightening of ICC power but as a new blend of the competing forms of regulation desired by shippers on the one hand and the railroads and their scientific management allies on the other.

All of this is in the literature. And what one would like to see in the paper is more awareness of what this scholarship has revealed about the forces precipitating and shaping regulatory growth and reform. Such an awareness, I think, would force changes in the paper's conclusions and in its suggestions of hypotheses in need of testing. In addition, one would like to see some discussion of how current makers of regulatory policy would benefit from a fuller understanding of the historical record. Unlike the other papers, it has little to say about the relevance of its historical findings for current assumptions and policy assessments.

MICHAEL E. LEVINE, California Institute of Technology and University of Southern California Law Center: The central question of public policy choice is "can we make it better?," rather than "can we make it optimal?" It is crucial to the analysis to determine "better than what?"

Since the world is imperfect, comparing perfect regulation with imperfect competition isn't going to help us much. Yet this is what John Panzar's paper does.

It does this by positing and then attacking a straw man—the proposition that without direct government intervention, the airline industry will work perfectly—and then presenting three models of situations in which nonregulation might lead to suboptimal results. The paper then specifies ways in which traditional or "evolved" regulatory techniques might, if perfectly applied, result in "socially optimal prices."

Greatly reduced regulation of airline markets seems to be producing results more acceptable than those that could be projected for even "enlightened" regulation. Panzar has not dealt with these facts and has not justified the particular specification of his models in the face of plausible alternatives that obviate their consequences. Theoretical analysis predicted that deregulation would produce more intense resource use (higher load factors) and fares closer to the costs of providing consumer-preferred levels of service (lower fares in many instances, more service at higher fares in some others) than had the previous regime of entry and price regulation. The facts suggest that the impact of administrative and statutory deregulation has been positive: Load factors are up. Average prices declined in nominal terms for over two years and in-

creased very slightly in nominal terms since 1977 in the face of substantial general inflation and enormous rises in the costs of airline input factors. Significant changes in the output configurations of the firms in the industry have accompanied improvements in aggregate productivity, suggesting that the previous output patterns under regulation were less efficient than practically possible.

Panzar refers to passenger pricing problems that a competitive bidding system would have engendered in 1935 even if the legislative architects had avoided the venality and folly of the 1930 and 1934 efforts. But if the government simply wanted the mail carried and regarded passenger service as a desirable nonessential by-product, there would be no harm in maximizing revenues from passengers, especially if passenger service could not generate rents. There is no short-term economic reason to prefer a scheme that recovers 40 percent of joint costs from passengers to one that recovers 50 percent. If there *are* economic or political reasons to favor passenger development over mail service development, those reasons could easily be accomodated through a bid qualification setting a maximum fare per mile designed to stimulate passenger traffic, making up the rest of the output costs through mail payments and obviating any need to consider entry regulation.

Because airline markets can now support more than one carrier on most routes, Panzar's prediction of unstable natural monopolies is of little relevance. Also, Panzar seems to limit the set of strategic responses by the incumbent monopolist to the choice of a single price at which it will produce its previous output. A real-world incumbent, however, probably has a less costly path of adjustment to the cost-minimized output than does an entrant building from zero. The incumbent could rely on this, advertising to potential entrants its strategies and the advantages of incumbency to discourage entry. Panzar seems to assume that the regulator will not allow a contraction of output by the incumbent, but does not explain why this same regulator would not require any new entrant to provide for production of the "lost" output once the incumbent threatens abandonment. Why will the entrant firm be allowed by regulators to earn rents when the incumbent firm is not? Whether Panzar's analysis has practical implications depends on his defense of his assumptions.

Profitability and Capital Costs for Manufacturing Corporations and All Nonfinancial Corporations

By DANIEL M. HOLLAND AND STEWART C. MYERS*

This paper is a progress report on our effort to measure and analyze the profitability of *U.S.* corporations over the postwar period. We are particularly interested in comparing the behavior of manufacturing corporations (*MC*) to that of the larger aggregate of all nonfinancial corporations (*NFC*). We analyzed *NFC*s in an earlier paper using data from both capital markets and the National Income and Product Accounts (*NIPA*). This paper's approach is basically the same.[1]

Rates of return derived from the *NIPA* are solid, useful numbers. However, we believe the capital market data are essential. A decline in the real rate of return on capital does not make investors worse off or weaken the incentive to invest if the real cost of capital declines proportionally. Also, an independent measure of rate of return can be obtained from returns realized by investors in debt and equity securities. If the only object were to measure corporate profitability, we believe market rates of return would be superior to the *NIPA* figures. However, both types of data are needed to investigate the determinants of profitability and to assess profitability relative to the cost of capital.

I. Data and Assumptions

Our results appear in Tables 1 and 2. We will briefly describe the statistics in these tables before attempting to comment or to interpret them.

The rates of return (*ROC*s) in Table 1 are the ratios of before- or after-tax operating income to capital stock. Capital stock is defined as the current value of plant, equipment, and inventories. Operating income includes interest, excludes inventory profit, and is calculated after depreciation of the net replacement cost of the capital stock.[2]

The *ROC*s for *NFC*s can be derived directly from the *NIPA*. The derivation for *MC*s is more difficult, however, and the estimates less satisfactory. Capital stock estimates for *MC*s are given in the *NIPA* on an establishment basis, but several items necessary for estimating operating income are available on a company basis only. The inventory valuation adjustment (*IVA*) is given on an establishment basis, but for all manufacturing, not just corporate manufacturing. A capital consumption adjustment for manufacturing is not available in the *NIPA*. The procedures we used to get around these difficulties are briefly described in the Appendix.

The returns to investors shown in Table 1 are the estimated real rates of return on a portfolio of all outstanding debt and equity securities issued by *MC*s or *NFC*s.[3] That is,

*Professors of finance, Massachusetts Institute of Technology. We wish to thank John Gorman of the Bureau of Economic Analysis, Bernard Horn, and Jerome Dausman.

[1]However, the *NFC* statistics given in this paper may differ from those presented in our earlier paper. The differences reflect data revisions and minor procedural changes necessary to ensure comparability of the *NFC* and *MC* series.

[2]Strictly speaking, operating income equals real income only in the absence of real holding gains on inventory and plant and equipment. In measuring *ROC*s we assume, in effect, that nominal holding gains are just sufficient to offset changes in the *CPI*. We checked this assumption in our earlier study, and found it to be true *on average* for *NFC*s, although there are substantial year-to-year variations which we think reflect difficulties in measuring short-term rates of asset appreciation from accounting data.

[3]Nominal rates were estimated first, and then converted to real rates by subtracting December to December percentage changes in the *CPI*.

TABLE 1—ESTIMATED REAL RATES OF RETURN FOR MANUFACTURING CORPORATIONS AND ALL NONFINANCIAL CORPORATIONS, 1946–78
(Shown in Percent)

	Manufacturing				All Nonfinancial			
Year	Return to Investors	After-Tax Return on Capital	Effective Tax Rate	Before-Tax Return on Capital	Return to Investors	After-Tax Return on Capital	Effective Tax Rate	Before-Tax Return on Capital
1947	−0.6	9.4	55.7	16.8	−5.1	7.0	52.1	14.6
1948	2.5	9.6	44.6	17.7	2.2	9.0	44.2	16.1
1949	18.3	10.1	37.9	16.1	17.1	8.4	38.7	13.7
1950	28.7	8.3	58.2	20.1	17.7	7.2	55.4	16.1
1951	20.9	7.7	63.4	21.0	13.0	6.3	61.5	16.2
1952	16.9	7.0	57.7	16.5	14.7	5.9	56.5	13.6
1953	−2.5	6.3	61.3	16.1	−0.7	5.2	59.3	12.7
1954	56.9	6.6	52.9	14.0	42.6	5.7	51.7	11.9
1955	34.9	8.7	52.8	18.5	24.4	7.3	50.8	14.8
1956	4.7	7.0	53.9	15.2	1.9	5.9	53.2	12.7
1957	−13.5	6.5	52.4	13.7	−10.6	6.0	51.3	11.5
1958	39.6	5.1	51.0	10.4	33.8	5.0	49.4	9.8
1959	10.8	7.8	49.5	15.5	8.3	6.4	48.4	12.4
1960	−2.6	7.2	48.5	13.9	0.3	6.1	46.9	11.4
1961	25.2	6.8	49.5	13.4	22.1	6.0	47.2	11.4
1962	−10.5	8.9	44.8	15.7	−7.3	7.7	41.7	13.1
1963	21.2	9.5	45.0	17.3	16.7	8.2	41.6	14.0
1964	14.8	10.8	42.1	18.8	13.1	9.3	38.7	15.2
1965	10.6	12.6	41.4	21.5	8.2	10.3	37.6	16.5
1966	−13.1	12.5	41.3	21.4	−11.6	10.2	37.5	16.3
1967	20.9	10.7	39.5	17.6	14.7	9.1	36.5	14.3
1968	5.2	9.7	44.8	17.6	4.7	8.3	40.9	14.1
1969	−13.6	7.8	46.1	14.5	−14.5	7.2	41.9	12.4
1970	−0.1	5.8	43.1	10.1	1.7	5.9	39.8	9.8
1971	11.0	6.3	43.6	11.2	10.1	6.2	39.0	10.1
1972	14.5	7.8	41.5	13.4	12.4	7.1	36.8	11.2
1973	−20.8	7.0	46.3	13.1	−19.7	6.5	40.0	10.9
1974	−31.4	2.7	66.4	8.0	−31.9	4.3	47.9	8.2
1975	22.8	5.3	39.7	8.7	21.4	5.5	37.7	8.8
1976	16.9	6.3	43.7	11.2	17.4	5.8	40.0	9.7
1977	−12.9	6.7	43.8	12.0	−11.7	6.1	39.8	10.1
1978	−3.3	6.7	44.0	11.9	−4.7	6.0	41.0	10.2
1947–50	12.2	9.4	49.1	17.7	8.0	7.9	47.6	15.1
1951–54	23.1	6.9	58.8	16.9	17.4	5.8	57.2	13.6
1955–58	16.4	6.8	52.5	14.5	12.4	6.1	51.2	12.2
1959–62	5.7	7.7	48.1	14.6	5.9	6.6	46.1	12.1
1963–66	8.4	11.4	42.5	19.8	6.6	9.5	38.9	15.5
1967–70	3.1	8.5	43.4	15.0	1.7	7.6	39.8	12.7
1971–74	−6.7	6.0	49.5	11.4	−7.3	6.0	40.9	10.1
1975–78	5.9	6.3	42.8	11.0	5.6	5.9	39.6	9.7
			Average for the Period 1947–78					
Mean	8.5	7.9	48.3	15.1	6.3	6.9	45.2	12.6
Standard Deviation	18.8	2.1	7.3	3.6	15.8	1.5	7.2	2.4

Table 2—Additional Statistics

	Manufacturing					All Nonfinancial			
Year	q	Capital-ization Rate[a]	Standard Deviation[b]	Adjusted Standard Deviation[b]	Market Debt Ratio, Midyear	q	Capital-ization Rate[a]	Standard Deviation[b]	Market Debt Ratio, Midyear
1947	.96	7.8	4.9	4.6	−.01	1.00	7.0	4.2	.16
1948	.80	12.3	4.8	4.6	−.01	.87	10.4	3.9	.18
1949	.60	16.8	4.9	4.7	−.01	.71	11.8	3.8	.23
1950	.74	11.3	3.3	3.1	−.06	.79	9.2	2.4	.17
1951	.62	12.5	3.6	3.4	−.04	.72	8.6	2.7	.19
1952	.60	11.6	3.7	3.5	.00	.72	8.2	2.8	.20
1953	.62	10.0	3.2	3.1	.00	.71	7.3	2.5	.18
1954	.69	9.5	4.0	3.8	.02	.77	7.4	3.0	.22
1955	.98	9.0	3.9	3.8	−.01	.97	7.5	2.9	.17
1956	.97	7.1	4.3	4.1	.00	.98	5.6	3.3	.16
1957	.92	7.1	4.1	3.9	.02	.92	6.1	3.2	.16
1958	.83	6.2	3.5	3.3	.05	.91	5.5	2.8	.19
1959	1.19	6.6	2.3	2.2	.03	1.15	5.6	2.0	.16
1960	1.15	6.2	3.1	3.0	.03	1.10	5.5	2.6	.16
1961	1.33	5.1	3.3	3.1	.04	1.29	4.7	2.8	.16
1962	1.31	6.7	4.5	4.3	.04	1.24	6.2	3.7	.19
1963	1.48	6.4	4.5	4.3	.04	1.39	5.9	3.7	.18
1964	1.73	6.3	2.2	2.1	.04	1.49	6.2	1.8	.18
1965	1.98	6.4	2.0	1.9	.05	1.57	6.5	1.7	.18
1966	1.66	7.5	2.9	2.8	.08	1.43	7.2	2.6	.18
1967	1.57	6.8	3.5	3.4	.10	1.41	6.5	3.2	.19
1968	1.68	5.8	3.5	3.4	.09	1.38	6.0	3.1	.17
1969	1.50	5.2	3.5	3.3	.12	1.31	5.5	3.1	.20
1970	1.01	5.7	4.3	4.1	.19	.97	6.0	4.1	.27
1971	1.21	5.2	4.4	4.2	.16	1.12	5.5	4.1	.24
1972	1.29	6.0	2.7	2.6	.16	1.20	5.9	2.5	.23
1973	1.10	6.4	2.8	2.7	.21	1.16	5.6	2.7	.26
1974	.54	4.9	4.3	3.7	.43	.92	4.7	4.9	.32
1975	.65	8.3	5.3	4.7	.22	.79	6.9	5.7	.30
1976	.68	9.3	3.9	3.7	.21	.88	6.6	3.6	.26
1977	.68	10.0	3.1	2.9	.25	.79	7.7	2.9	.31
1978	.56	12.0	3.1	2.9	.25	.71	8.5	2.8	.35

[a]Shown in percent.
[b]Shown in percent per month.

it is a weighted average of debt and equity returns, where the weights are the ratios of debt to firm value and equity to firm value. These are market, not book, ratios. We estimated the weights by capitalizing interest and dividend flows, following the procedures described in our earlier paper.

Our major reservation about the capital market data shown in Tables 1 and 2 concerns the equity indexes underlying them. The estimates for *NFC*s and *MC*s reflect dividend yields and returns for the Standard and Poor's Composite and Industrial Indexes, respectively. Returns on these indexes are highly correlated, and either one is likely to pick up major shifts in market values over time. But *if* there are significant differences between *MC* and *NFC* equities, we doubt that these two indexes will pick them up. We plan to construct indexes designed to match the *MC* and *NFC* sectors.

The statistics in Table 2 require only brief comment. The variable q is the ratio of market value to capital stock; in other words, the ratio of the value of all tangible and intangible assets to the measured value of tangible assets. The absolute level of q ought to be taken with a grain of salt; but

changes in q reflect changes in investors' forecasts of profitability relative to the cost of capital.[4]

The standard deviations are based on twenty-four monthly observations of real rates of return to investors. The adjusted standard deviation for *MC*s is the unadjusted figure multiplied by the correlation coefficient between the rates of return to *MC* and *NFC* investors. This adjustment removes that part of the *MC* business risk that would be eliminated in a diversified portfolio of all *NFC* securities.

The capitalization rates are the ratios of operating income to market value. The debt ratio is the ratio of debt to debt plus equity at market value.

II. Comparing *MC*s and *NFC*s

The similarities between the time-series for *MC*s and *NFC*s are more striking than the differences. We will discuss the similarities first.

Both *MC*s and *NFC*s have fared poorly since the mid-1960's, which with hindsight looks more and more like a vigorous golden age. Average real rates of return to investors were actually negative over the past ten years: -1.7 percent for *MC*s and -1.95 percent for *NFC*s. Before and after tax, *ROC*s have fallen substantially. So has q, which confirms that profitability has fallen relative to the real cost of capital.

Thus someone with a ten- or fifteen-year memory would see a steep downtrend. However, the trend disappears when you look back further and correct for changing business conditions. Table 3 displays the results of regressing the *ROC* series on time, the percentage change in real *GNP*, and the percentage change in the *CPI*. There is a hint of a downward trend in the before-tax *ROC*s, but none in the after-tax figures. (The reason for this difference is that the effective tax rate on real operating income has fallen over the postwar period.)

The rate of inflation affects only after-tax profitability. In the short run at least, increased inflation increases the effective tax rate. However, over longer periods, the effective tax rate has decreased in the face of substantial inflation. The major reason is that inflation and increasing debt ratios have increased interest payments and shielded a greater proportion of operating income from the corporate tax. Modifications of the tax, for example, accelerated depreciation, shortening of depreciable lives and the investment tax credit, also contributed to the decline.

The capitalization rates can be thought of as earnings-price ratios generalized to include both debt and equity, with income stated in real terms. The most interesting feature of the capitalization rate series is its stability, at least from the mid-1950's through 1976. In other words, changes in q over this period have been primarily due to changes in real profitability (*ROC*), and apparently not to shifts in the real cost of capital. (We cannot be sure the real cost of capital has been stable, however, because the capitalization rate is only a rough measure of it.) However, in the last two years the capitalization rate has moved sharply upwards, to 8.5 for *NFC*s and 12.0 for *MC*s in 1978, the highest rate since the late 1940's. We cannot explain this shift; it may prove transitory.

The standard deviation series measure business risk. Here we see no evident trend. We can only note the quiet period in the mid-1960's and the extreme volatility of the mid-1970's. Note that the standard deviations have come down since the 1975 peak, so high business risk does not explain the high recent capitalization rates.

Now we turn to the differences between *MC*s and *NFC*s. The most obvious difference is that the *MC*s seem to have been more profitable. The before-tax *ROC*s are consistently higher for *MC*s than for the

[4]Market value MV equals the long-run earnings from assets already in place Y capitalized at the cost of capital ρ, plus the present value of future growth opportunities (*PVGO*): $MV = Y/\rho + PVGO$; Y equals *ROC* times the capital stock *CS*. Thus $q = MV/CS = ROC/\rho + PVGO/CS$. An increase in *ROC* relative to ρ increases q directly, and also indirectly through *PVGO*. Higher future profitability from existing assets usually means that future investment is more profitable also.

TABLE 3—REGRESSION ANALYSIS OF REAL RATE OF RETURN ON CAPITAL, 1947–78

Series	Time	Percent Change in Real *GNP*	Percent Change in *CPI*	R^2
Manufacturing, Before Tax	−.20	.60	.06	.87
	(−1.30)	(8.36)	(.60)	
*NFC*s, Before Tax	−.13	.35	−.04	.90
	(−1.20)	(8.47)	(.60)	
Manufacturing, After Tax	.09	.25	−.29	.77
	(1.12)	(4.32)	(−3.51)	
*NFC*s, After Tax	.08	.17	−.18	.81
	(.96)	(4.58)	(−3.54)	

Note: Fitted by standard Cochrane-Orcutt procedure; *t*-statistics shown in parentheses.

larger *NFC* aggregate. The average gap between the two peaked at 4.3 percentage points from 1963–66, although it has declined to an average of 1.3 percentage points since then. The gap also exists in after-tax *ROC*s, at least through the mid-1960's. Since then the after-tax *ROC*s have, on the average, nearly converged. The convergence of after-tax *ROC*s reflects growing differences in effective tax rates, which have been consistently lower for *NFC*s.

The higher profitability of *MC*s also shows up in higher average rates of return to investors and in generally higher *q*s.

One possible explanation for higher return is higher risk. The evidence we have favors that explanation. We note that the *MC*s' standard deviations are a bit higher than the *NFC*s', except in the 1970's.[5] The *MC*s' before- and after-tax *ROC*s are more variable, and more sensitive to the rate of growth of real *GNP* and the *CPI*. (See Table 3.) *MC*s waxed fatter in the mid-1960's and suffered more in the 1974 crunch. The *MC*s' lower debt ratios may be lenders' and managers' response to higher business risk. Finally, capitalization rates for *MC*s are generally *higher* than for *NFC*s. Greater risk ought to mean higher capitalization rates, other things equal.

III. Conclusion

We have emphasized that the estimates we have derived for *MC*s are probably less accurate than the *NFC* statistics. Accordingly we have restricted our comments to general points that are unlikely to be upset by more detailed analysis or improved measurement procedures.

The main points are these. Both *MC*s and *NFC*s show the same behavior over time:[6] great performance in the 1960's, much poorer performance since, but no longer-term trends. Business risk and the real cost of capital seem to have been stable over most of the postwar period. *MC*s appear to be riskier than the larger *NFC* aggregate, however, and *MC*s' average real rates of return have been higher.

APPENDIX

The estimates of the return on capital in Table 1 are derived from data in U.S. Bureau of Economic Analysis, 1976b, supplemented by annual revision and updating in the *Survey of Current Business*; *BEA*, 1976a, and unpublished series generously provided by John Gorman of *BEA*.

For both *NFC*s and *MC*s the capital stock equals the average of beginning- and

[5]We do not put much weight on this evidence alone, because we use indexes of equity returns that are highly correlated, and may miss significant differences between the *MC* and *NFC* sectors. See Section I of this paper.

[6]This is not preordained. The *MC* capital stock was 54 percent of *NFC* stock in 1947 and only 35 percent in 1978.

end-of-year values of plant and equipment (*BEA*, 1976a), plus second-quarter inventories (*BEA* unpublished data). (Prior to 1958, inventories, too, are averages of beginning- and end-of-year values.)

In what follows all numbered references are to tables in *BEA* (1976b) and the *Survey of Current Business*.

For *NFC*s all the elements of the numerator—corporate profits with inventory valuation and capital consumption adjustments, net interest, and profits tax liability—are taken directly from (1.15).

For *MC*s however, a number of items were estimated. While we found the methods developed by Jerome Dausman very helpful, a number of loose ends remain. As a consequence, the *ROC*s for *MC*s are not as accurate as for *NFC*s, and are subject to revision.

To estimate the capital consumption adjustment for corporate manufacturing, we applied the ratio of capital consumption allowances with and without the capital consumption adjustment for all corporations (8.7) to capital consumption allowance for all manufacturing (6.1), and then subtracted corporate manufacturing capital consumption allowance (derived by subtracting noncorporate manufacturing capital consumption allowance (6.15) from manufacturing capital consumption allowance (6.1)) to end up with corporate manufacturing capital consumption adjustment. In effect, we assumed that for manufacturing corporations the net capital stock was the same age as for all corporations.

Taxable profit on an establishment basis was estimated by adding back the *IVA* for all manufacturing (5.8) to all manufacturing profit-type income (6.1) and subtracting manufacturing sole proprietorship and partnership income (6.14).

Corporate manufacturing profits tax liability on an establishment basis was estimated by using the effective tax rate on a company basis (6.19 and 6.20).

Up to 1970, we used the *IVA* of (6.16) which is for corporate manufacturing on a company basis and "differs from that which adjusts book value inventories." From 1971 on, our *IVA* was that fraction of the manufacturing *IVA* of (5.8)—which is not the same as that "which adjusts business income" in (6.16)—that the corporate manufacturing *IVA* (6.16) is of the corporate and noncorporate manufacturing *IVA* (6.16). Our *IVA* is not a good estimate, and could be seriously off in 1974 and 1978, when the *IVA* was extremely high, causing our estimates of the *ROC* to be too low.

Net interest (6.17) is for all manufacturing, and, therefore, an overstatement for corporate manufacturing. However, the error is so slight—*ROC*s would be off by less than one-tenth of one percentage point—that we made no adjustment on this score.

REFERENCES

J. F. Dausman, "Corporate Manufacturing Rates of Return, 1947–76," unpublished Master of Science dissertation, M.I.T. 1978.

D. M. Holland and S. C. Myers, "Trends in Corporate Profitability and Capital Costs," in Robert Lindsay, ed., *The Nation's Capital Needs: Three Studies*, Washington 1979.

U.S. Bureau of Economic Analysis, (1976a) *Fixed Nonresidential Business and Residential Capital in the United States, 1925–75*, Washington 1976.

———, (1976b) *The National Income and Product Accounts of the United States, 1929–74*, Washington 1976.

U.S. Office of Business Economics, *Surv. Curr. Bus.*, Washington, various issues.

Rates of Return by Industrial Sector in the United States, 1948–76

By BARBARA M. FRAUMENI AND DALE W. JORGENSON*

The purpose of this paper is to present estimates of nominal and own-rates of return for forty-six industrial sectors of the *U.S.* economy. These estimates are derived within a system of national accounts that includes measures of prices and quantities of output and input for individual industrial sectors, and prices and quantities of income and expenditure, saving and investment, and wealth for the *U.S.* economy as a whole.[1] These prices and quantities are combined to provide measures for each of the accounting concepts that are adjusted for inflation.[2]

A system of prices for capital goods is an essential component of a complete system of national accounts. The prices of new capital goods are compared with the prices of consumption goods and services in determining the allocation of the national product between investment and consumption. The prices of the services of capital goods currently in existence are compared with the prices of labor services in selecting the relative proportions of capital and labor services in production.

In the absence of taxation, the price of capital services, say $p_{K,t}$, can be expressed in terms of the price of capital goods $p_{A,t}$, the rate of return r_t, depreciation $p_{D,t}$, and revaluation $p_{A,t} - p_{A,t-1}$:

$$p_{K,t} = p_{A,t-1} r_t + p_{D,t} - (p_{A,t} - p_{A,t-1})$$

This relationship results from representing the price of acquisition of capital goods as the discounted value of future rentals.

Outlay on capital services can be expressed as the sum of outlays on all types of capital services. Each outlay is the product of the price of capital services and the corresponding quantity of capital, say K_{t-1}. If there were only one asset, property compensation could be represented in the form:

$$p_{K,t} K_{t-1} = p_{A,t-1} r_t K_{t-1} + p_{Dt} K_{t-1} - (p_{A,t} - p_{A,t-1}) K_{t-1}$$

Given data on property compensation, the acquisition price of capital, and depreciation, this expression determines the rate of return r_t, and therefore the price of capital services. If there are several assets, property compensation is the sum of the value of capital services for all assets.

Again, in the absence of taxation, the rate of return can be expressed in the form:

$$r_t = \frac{p_{K,t} K_{t-1} - p_{D,t} K_{t-1} + (p_{A,t} - p_{A,t-1}) K_{t-1}}{p_{A,t-1} K_{t-1}}$$

The rate of return is equal to the ratio of property compensation, less depreciation and plus revaluation, to the value of capital at the beginning of the period. This is a *nominal* rate of return, since it includes the revaluation of assets. The corresponding *own*-rate of return excludes revaluation and takes the form:

$$r_t - \frac{p_{A,t} - p_{A,t-1}}{p_{A,t-1}} = \frac{p_{K,t} K_{t-1} - p_{D,t} K_{t-1}}{p_{A,t-1} K_{t-1}}$$

*Boston College and Harvard University, respectively.

[1]The system of national accounts was originally designed and implemented for the private sector of the *U.S.* economy by Laurits Christensen and Jorgenson (1969, 1970, 1973a,b). This system has been extended to incorporate production accounts for individual industrial sectors of the *U.S.* economy by Fraumeni and Jorgenson.

[2]Adjustments of a system of national accounts for inflation are discussed by Phillip Cagan and Robert Lipsey. Adjustments of business accounts for inflation have been discussed from an economic viewpoint by Solomon Fabricant, and by John Shoven and Jeremy Bulow (1975, 1976).

TABLE 1—SECTORAL NOMINAL AND OWN-RATES OF RETURN, ANNUAL AVERAGES, 1948–76

Industry	Nominal	Own	Industry	Nominal	Own
Agriculture	0.0749	0.0350	Electrical Machinery	0.1212	0.0825
Agricultural Services	0.0693	0.0388	Transportation Eqpt.	0.0535	0.0127
Metal Mining	0.0900	0.0462	Motor Vehicles	0.2846	0.2450
Coal Mining	0.1424	0.1021	Professional Photographic Eqpt.	0.1420	0.1052
Crude Petro.	0.1240	0.0812	Miscellaneous Manufacturing	0.1382	0.1030
Nonmetalurgical Mining	0.1522	0.1135	Railroads	0.0747	0.0346
Construction	0.1471	0.1078	Street Rail, Bus	0.1712	0.1352
Food	0.1031	0.0667	Trucking Services	0.1435	0.1031
Tobacco	0.1350	0.1085	Water Transportation	0.0727	0.0455
Textiles	0.0903	0.0577	Air Transportation	0.0199	−0.0152
Apparel	0.1052	0.0809	Pipelines	0.1106	0.0697
Paper	0.1283	0.0864	Transportation Services	0.0983	0.0632
Printing & Publishing	0.1069	0.0654	Tel. & Tel.	0.1460	0.1126
Chemicals	0.1322	0.0950	Radio & Tel.	0.1514	0.1188
Petroleum Refining	0.1240	0.0812	Electric Utilities	0.1306	0.0836
Rubber	0.1052	0.0660	Gas Utilities	0.1454	0.0988
Leather	0.0980	0.0724	Water	0.1824	0.1399
Lumber & Wood	0.2045	0.1648	Wholesale Trade	0.1269	0.0936
Furniture	0.1137	0.0741	Retail Trade	0.1017	0.0677
Stone, Clay, Glass	0.1123	0.0738	Finance, Insurance & Real Estate	0.0560	0.0178
Primary Metal	0.0900	0.0462	Services	0.0913	0.0533
Fabricated Metal	0.1051	0.0595	Households	0.0793	0.0442
Machinery excluding Electrical	0.1460	0.1068	Institutions	0.0793	0.0309

If rates of growth of asset prices differ among assets, the own-rates of return that enter into the prices of capital services are different for each asset.

In the presence of taxation, the value of capital services must be reduced by tax liabilities in determining the rate of return. The nominal rate of return for all assets is equal to the ratio of property compensation, less depreciation and tax liabilities and plus revaluations, to the value of capital at the beginning of the period. The own-rate of return for all assets excludes revaluation, as before.

We combine data on property compensation and the value of capital for each industrial sector with information on the tax structure for property compensation and estimates of depreciation and revaluation of existing stocks of capital to obtain estimates of rates of return for all forty-six industrial sectors. In Table 1 we present nominal rates of return for each sector, including earnings on capital assets and gains from revaluation. We also present own-rates of return for each sector, including only earnings on assets.

Nominal rates of return for the period 1948–76 vary from 0.0199 in air transportation to 0.2846 for motor vehicles and equipment. Thirty of the forty-six sectors have rates of return between 9 and 15 percent. Transportation equipment, except for motor vehicles, and ordnance has a nominal rate of return for the period as a whole of 0.0535; finance, insurance, and real estate has a nominal rate of return of 0.0560. By contrast lumber and wood products has a nominal rate of return of 0.2045, and water supply and sanitary services has a nominal rate of return of 0.1824 percent.

Own-rates of return for the period as a whole vary from a negative 0.0152 for air transportation to 0.2450 for motor vehicles and equipment. Thirty-nine of the forty-six sectors have own-rates of return between 3 and 12 percent. Transportation equipment, except for motor vehicles, and ordnance has an own-rate of return of 0.0127 for the period as a whole, while finance, insurance, and real estate has an own-rate of return of 0.0178. By contrast, lumber and wool products has an own-rate of return of 0.1648,

TABLE 2—CLASSIFICATION OF ANNUAL AVERAGES FOR SECTORAL NOMINAL RATES OF RETURN BY PERIOD, 1948–76

Average Rate of Growth in Percent	1948–76	1948–52	1953–56	1957–59	1960–65	1966–68	1969–72	1973–76
Less than 3	1	1	4	3	1	0	1	0
3 to less than 6	2	4	8	9	6	1	1	2
6 to less than 9	6	12	10	11	19	3	4	1
9 to less than 12	15	7	13	15	10	18	19	3
12 to less than 15	16	12	8	6	6	10	10	14
15 to less than 18	3	4	0	2	3	7	2	11
18 to less than 21	2	2	1	0	0	3	3	5
21 or more	1	4	2	0	1	4	6	10

TABLE 3—CLASSIFICATION OF ANNUAL AVERAGES FOR SECTORAL OWN-RATES BY PERIOD, 1948–76

Average Rate of Growth in Percent	1948–76	1948–52	1953–56	1957–59	1960–65	1966–68	1969–72	1973–76
Less than 0	1	3	0	1	1	0	1	2
0 to less than 3	2	9	9	8	2	2	5	8
3 to less than 6	11	9	12	10	7	6	15	11
6 to less than 9	15	10	11	15	19	16	12	11
9 to less than 12	13	5	11	9	9	8	4	6
12 to less than 15	2	5	0	3	4	7	2	4
15 to less than 18	1	2	1	0	3	4	3	0
18 or more	1	3	2	0	1	3	4	4

and water supply and sanitary services has an own-rate of return of 0.1399.

In Table 2 we present a distribution of nominal rates of return for the forty-six industrial sectors included in our study. In Table 3 we present a distribution of own-rates of return for these same industrial sectors. To analyze the patterns of change in these distributions across subperiods of time, we have divided the period 1948–76 into seven subperiods corresponding to the years between cyclical peaks of the *U.S.* economy in 1948, 1953, 1957, 1960, 1966, 1969, and 1973. Each subperiod, except for the last, begins with a cyclical peak and runs through the year prior to the following cyclical peak. The last subperiod begins with the peak year of 1973, but includes only the years 1973–76.

In comparing rates of return across subperiods of time, we find that the distribution of average nominal rates of return has shifted upward between the 1950's and 1960's, and between the 1960's and the 1970's, reflecting the increasing rate of inflation in the prices of capital goods between these decades. By contrast, the distribution of average own-rates of return, which excludes gains from the revaluation of assets, does not exhibit a similar trend. This distribution appears to be relatively stable over time, reflecting the stability of earnings on assets.

The aggregate nominal rate of return for the period 1948–76 is 0.0848; for the seven subperiods included in Table 2 the aggregate nominal rates of return of 0.0859 for 1947–52, 0.0577 for 1953–56, 0.0563 for 1957–59, 0.0647 for 1960–65, 0.1011 for 1966–68, 0.1038 for 1969–72, and 0.1309 for 1973–76. The distributions of nominal rates of return across industries reflect the aggregate rates. The distributions are concentrated at relatively low levels for the subperiods 1953–56, 1957–59, and 1960–65. Distributions for the remaining subperiods are higher, reflecting higher rates of inflation in asset prices.

The aggregate own-rate of return for the period 1948–76 is 0.0478; for the seven subperiods the aggregate own-rates of return are 0.0432 for 1948–52, 0.0414 for 1953–56,

0.0313 for 1957–59, 0.0515 for 1960–65, 0.0605 for 1966–68, 0.0514 for 1969–72, and 0.0484 for 1973–76. Since revaluations are excluded from the own-rates of return, the own-rates of return show much less variation over time than the nominal rates of return. The subperiods 1960–65, 1966–68, and 1969–72 are characterized by high average own-rates of return. The periods 1948–52, 1953–56, and 1957–59 are characterized by low-average own-rates of return. The own-rate of return for the most recent subperiod, 1973–76, is almost identical to the average for the period as a whole.

Our first conclusion is that there are surprisingly large differences in rates of return among sectors. For every dollar of capital shifted from air transportation to motor vehicles and equipment, the *U.S* economy would have earned $.2647 annually for the period 1948–76. Similarly, for every dollar shifted from transportation equipment, other than motor vehicles, and ordnance, to lumber and wood products, the *U.S* economy would have earned $.1510 annually. The differences in marginal productivity of capital among sectors, as measured by the own-rates of return, are comparable to the differences in nominal rates of return. Differences in own-rates of return provide opportunities to increase output without increasing capital stock by reallocating capital among sectors.

Our second conclusion is that substantial differences in rates of return among sectors have persisted over the period 1948–76. If rates of return were distributed independently over time, we would expect the variances of the distributions to be inversely proportional to the number of time periods. Our final conclusion is that own-rates of return for the most recent subperiod, 1973–76, are about average by postwar standards, while rates of return for the period 1966–68 were exceptionally high, and rates of return for the period 1957–59 were exceptionally low. These results are consistent with the findings of six studies of rates of return for all nonfarm, nonfinancial corporations surveyed by Cagan and Lipsey and with a recent study of rates of return for 187 individual corporations by William Brainard, Shoven, and Leonard Weiss.[3]

Our objective in measuring rates of return has been to account for the role of capital in explaining postwar *U.S.* economic growth.[4] Our principal findings can be briefly summarized as follows: The growth of capital stock accounts for 31 percent of the growth of output for the *U.S.* economy as a whole over the period 1948–76. The reallocation of capital among classes of assets within industries accounts for another 15 percent of the growth of output. Finally, the reallocation of capital among industries has contributed only a little more than 2 percent of the growth of output for the period 1948–76. Our overall conclusion is that differences in rates of return among sectors have failed to bring about a reallocation of capital that has contributed substantially to the growth of output.

Our estimates of rates of return by industrial sector can be combined with other data to analyze the allocation of capital within the *U.S.* economy in greater detail. By combining our estimates of rates of return with a financial measure of the cost of capital, it would be possible to analyze the impact of rates of return on investment and the market valuation of financial claims on assets. Similarly, by combining our estimates with the underlying effective tax rates by sector, it would be possible to apply our results to an analysis of the impact of the taxation of income from capital on capital allocation. Finally, comparisons of rates of returns across industries could be employed in the study of the impact of industrial organization on the allocation of capital.

[3]Additional estimates of the rate of return have been given by Richard Kopcke, Michael Lovell, and William Nordhaus.

[4]The results of this study and a detailed description of the sources and methods employed for the estimates of rates of return presented in Tables 1, 2, and 3 are given by Fraumeni and Jorgenson.

REFERENCES

W. C. Brainard, J. B. Shoven, and L. Weiss,

"The Financial Valuation of the Return to Capital," *Brookings Papers*, forthcoming.

Phillip Cagan and Robert E. Lipsey, *The Financial Effects of Inflation*, Cambridge 1978.

L. R. Christensen and D. W. Jorgenson, "The Measurement of U.S. Real Capital Input, 1929–67," *Rev. Income Wealth*, Dec. 1969, *15*, 203–320.

______ **and** ______, "U.S. Real Product and Real Factor Input, 1929–1967," *Rev. Income Wealth*, Mar. 1970, *16*, 19–50.

______ **and** ______ (1973a), "U.S. Income, Saving, and Wealth, 1929–1969," *Rev. Income Wealth*, Mar. 1973, *16*, 19–50.

______ **and** ______ (1973b), "Measuring Economic Performance in the Private Sector," in Milton Moss, ed., *The Measurement of Economic and Social Performance*, New York 1973, 233–351.

S. Fabricant, "Accounting for Business Income under Inflation: Current Issues and Views in the United States," *Rev. Income Wealth*, Mar. 1978, *24*, 1–24.

B. M. Fraumeni and D. W. Jorgenson, "The Role of Capital in U.S. Economic Growth, 1948–1976," in George von Furstenberg, ed., *Capital, Efficiency, and Growth*, Cambridge, forthcoming.

R. Kopcke, "The Decline in Corporate Profitability," *New England Econ. Rev.*, May/June 1978, 36–60.

M. C. Lovell, "The Profit Picture: Trends and Cycles," *Brookings Papers*, Washington 1978, *3*, 769–89.

W. D. Nordhaus, "The Falling Share of Profits," *Brookings Papers*, Washington 1974, *1*, 169–208.

J. B. Shoven and J. I. Bulow, "Inflation Accounting and Nonfinancial Corporate Profits: Physical Assets," *Brookings Papers*, Washington 1975, *3*, 557–98.

______ **and** ______, "Inflation Accounting and Nonfinancial Corporate Profits: Financial Assets and Liabilities," *Brookings Papers*, Washington 1976, *1*, 15–57.

Vignettes on the World Capital Market

By ARNOLD C. HARBERGER*

This paper is motivated by a sense that we as a profession need to understand, much better than we now do, how the world capital market works. We seem to be genuinely schizophrenic in the ways we build models—many of them are closed-economy models in which the rest of the world does not even appear, yet others of them are models of the "small, open economy" in which hardly any degree of freedom is left for economic policy to influence events. Lying behind the schizophrenia is, I believe, a genuine ignorance on our collective part of how the world capital market works.

If there is really only one capital market linking most of the economies of the world (I do not want to enter here into how the Soviet and Chinese spheres of influence would fit in), then there is presumably something called a world interest rate, which would become a datum (or exogenous variable) for nearly all of them. A shift of the investment schedule in any such country would simply result in an inflow of funds from the rest of the world—not in a rise in interest rates. An increase in saving, likewise, would simply spill over the national boundary, and would not result in any change in local interest rates or investment.

If, on the other hand, there is little relevance to the concept of a world capital market, then one would expect interest rates in the different countries to be governed by internal factors, being sensitive to shifts in investment and savings, and presumably being influenced by the relative scarcity of capital within each country.

Martin Feldstein and I have each become linked with a particular (and probably polar) interpretation. I, looking at rates of return in different countries, and finding them basically uncorrelated with the capital-labor ratio, concluded that there must be some force operating to prevent such a correlation from emerging, and that the most natural explanation was that the world capital market was alive and well and living in Zurich.

Feldstein, drawing upon the arguments that in a very well-functioning world capital market a) shifts in a nation's saving propensity would not affect its investment rate, and b) shifts in a nation's investment schedule would not affect its savings rate, studied correlations between the savings and investment rates of different countries, and found them to be strongly positive.

From my experiments (1978) I drew the lesson not only that the rate of return to capital is brought into rough equalization through the international capital market, but also that the relative abundance or scarcity of labor (which does not move so readily across national boundaries) has a great deal to do with the determination of real wages. I found that the ratio of different nations' capital stocks to their quality-adjusted labor forces explained a) well under a third of the variation in their rates of return to capital, and b) around 90 percent of the variation in the rate of earnings of labor of given quality.

From their experiments, Feldstein and Charles Horioka conclude that "it is appropriate, at least as an approximation, to study income distribution in general and tax incidence in particular with models that ignore international capital mobility. [Also], the evidence implies that the national return on domestic saving is approximately equal to the pretax domestic marginal product of capital since such saving does increase the capital stock rather than either flowing abroad or replacing foreign investment at home." They found that, for the *OECD* countries in recent years, saving differs from investment, on the average by only a little more than 1 percent of *GDP* and by only

*University of Chicago.

around 5 percent of the component savings or investment figures themselves.

My hunch is that by the time we have given this subject something like the amount of attention it deserves, both Feldstein and I will have backed off from an extreme interpretation of our words. My guess is that the answer lies in some sort of middle ground, a point to which I will return at the end of this paper. The present paper represents an effort to move further toward a correct interpretation of how the world capital market works, adding a few more pieces of evidence, and trying to digest more fully what we already have.

I. More on Rates of Return

In my initial paper on this subject, I estimated national aggregate rates of real return to capital in eighteen countries. This was done using national accounts data, building up reproducible capital stocks separately for buildings, machinery and equipment, and inventories, and making specific assumptions to exclude land rents from the income figure in the numerator of the rate-of-return ratio. Apart from Korea, which was a distinct outlier in all experiments, the national aggregate rates of return of seven developed countries ranged between 4.4 percent for Sweden and 8.5 percent for the United States, while those for the less-developed countries (*LDCs*) ranged from 5.5 percent for Portugal to 10.8 percent for Jamaica. The corresponding ranges for the private after-tax rate of return were 3.1–7.6 percent for the developed countries and 3.9–10.2 percent for the *LDCs* (except Korea).

In interpreting these results it is important to be aware of what sort of question one is trying to answer. If there were a perfect capital market, why should there be any difference at all in the results, assuming we were measuring capital and its rate of return in ways that reflected the reality of the equalization process? My initial approach was concerned with the influence of different capital-labor ratios upon the rate of return; if we accept that as the right frame for the question, there is no doubt that the answer is as indicated—a weak influence if any, and negligible in magnitude. But that is not the only approach that can be taken.

Among the questions that arose in discussions of my original paper was one concerning the concept and measurement of the private after-tax rate of return. For better or worse, it included in the denominator the capital represented by the stock of residential housing, and in the numerator the actual and imputed rents attributed (in the national accounts) to that stock. Given a) the likely shakiness of rent imputations in many countries, b) the degree of government intervention through direct subsidies, tax exemptions, mortgage policies, etc. in the housing market just about everywhere, and c) the importance of housing as a component of the private capital stock, might one not get a clearer picture of the workings of the international capital market by simply excluding residential housing from the capital stock, and the corresponding actual and imputed rents from the numerator of the rate of return?

Giving an affirmative answer to the above question (in spite of some trepidations based on the importance of major financial institutions both in the mortgage market and in the obviously quite-perfect international market for liquid funds), I tried to adapt my earlier estimating procedure to produce estimates of the private rate of return to capital excluding housing. All in all, that was a frustrating exercise. Only for a few countries, mainly industrial ones, was one able to get data relative to rents, and even here the data were contaminated, including items like fuel, electricity, water, maintenance labor, etc. I abandoned the experiment when I found myself unable to make a very precise assumption about the proportion that those elements occupied in "total rent."

Faced with an inability to produce a set of results for different countries, based on the application of a single uniform procedure, I turned to a number of separate studies, done by different people and for different countries, but mainly following similar methodologies. The studies either directly measured the private rate of return to

TABLE 1—ESTIMATES OF THE PRIVATE RATE OF RETURN TO PRIVATE CAPITAL
(Return Defined Net of all Taxes Paid by Enterprises, Gross of Personal Income Taxes)

Country	Author	Data Base	Period	Estimated Rate of Return[j]
Argentina	Petrei[a]	Corporate Records	1961–67	10.24
Brazil	Langoni[b]	Corporate Records	1954–67	8.05[c]
Brazil	Novaes[d]	Corporate Records of foreign companies making direct investments in Brazil	1949–71	7.6
Canada	Jenkins[e]	Tax Statistics	1965–69	6.34
Colombia	Harberger[f]	National Accounts	1967	12.0
India	Harberger[g]	Corporate Records	1955–59	7.1
United States	Stigler[h]	Corporate Records	1947–57	7.78
United States	Jorgenson-Griliches[i]	National Accounts	1955–65	12.7

[a]A stratified sample of 600 enterprises; private rate of return calculated using double declining balance depreciation. See Amalio Petrei, Table 27, p. 70.

[b]Aggregate rate of return to physical capital based on surveys by Conjuntura Economica. The surveys cover some two-thirds of corporate investment in Brazil, and leave out public services and commerce. Data from Carlos Langoni, ch. 3.

[c]Eleven and one-half percent of return before corporation taxes, reduced by a 30 percent corporation tax.

[d]Cash flows from foreign investors to Brazil, minus net flows from Brazil to foreign investors, over twenty-one years. The 1949 stock of own-capital (estimated by perpetual inventory methods) was taken as the first investment, and the 1971 stock was taken as the last payment. See Rubem de Freitas Novaes, Table 4, p. 43.

[e]Average private-sector rate of return for "All Industries," 1965–69. See Glenn Jenkins, Table 6, p. 39.

[f]Returns to private-sector capital stock less housing (7.29 billion 1958 pesos) less taxes paid by corporations (.71 billion pesos) divided by beginning-of-year private-sector capital stock less housing (54.95 billion pesos). See my 1978 paper, Table 6.21, p. 153.

[g]Net income (Y_c) (see my 1972b paper, p. 216) adjusted for corporate tax provision (Reserve Bank of India, *Bulletin*), and expressed as a fraction of capital stock (deflated net assets, K_2) (see my 1972b paper, p. 218). Basic sample was a survey of 1,001 corporations by the Reserve Bank of India.

[h]The median of the annual rates of return reported for manufacturing industries for the years 1947–57 (see George Stigler, Table 10, col. 2, p. 35).

[i]The median of the annual implicit rates of return, after taxes, reported in Jorgenson and Griliches, Table VI, col. 4, p. 268.

[j]Shown in percent.

private capital, or else produced results from which I was able to take the subsequent steps to estimate that private rate of return. The results are presented in Table 1.

Certainly the general impression emerging from Table 1 is not inconsistent with the conclusions I had reached on the basis of my earlier work. There seems to be no strong and systematic tendency for rates of return to be high in countries with a low capital stock per worker. Indeed, the estimates by Dale Jorgenson and Zvi Griliches of the *U.S.* rate of return are the highest of all (these, by the way, are based on a different methodology, and include residential housing).

On the basis of these results one can speak only of a rough and general tendency for equalization of rate of return across countries. Capital's rewards are much more nearly equalized than labor's. The most direct and plausible explanation of this fact is that international capital movements are much more important and much less impeded than are international movements of labor.

II. More on International Capital Flows

The evidence in the Feldstein-Horioka paper is of three types: (a) that the difference between a country's saving and in-

vestment rates tends to be small, (b) that cross sectionally, there is a high correlation between savings and investment rates of different countries, and (c) that there is a tendency for investment and savings rates of a given country in different years to be positively correlated. Of these three, the evidence is more persuasive for (a) and (b) than for (c), which produces, as the authors note, high variability of regression coefficients across countries.

I think the underlying question is how far one can go in interpreting this evidence as somehow testing for the relevance of the international capital market. I would expect there to be a very low correlation between the savings and investment rate of, say, census tracts or city blocks. I would expect there to be a higher correlation between these rates for major cities, and a still higher correlation for states, and higher yet for major regions.

The intuition behind this expectation is that it is very easy for the investment in a city block to be many times the income of the block's residents (namely, in the construction of any large building), let alone their saving flow. Similarly, it is natural for savers to hold their savings in forms (savings accounts, corporate shares, etc.) that go to finance many investments unrelated to the block.

As one moves up to more aggregative levels, correlations between saving and investment rates will grow for at least two reasons. First, capital-to-income ratios will tend to be more similar across cities than, say, across blocks. If one thinks of net investment as providing for the growth of the capital stock and the net saving as being related to income, a cross-sectional relationship will presumably produce better and better correlations as the K/Y ratios of the individual observations become more homogeneous. Second, for the ratios of gross investment and gross saving, there is the obvious common component of capital consumption allowances, imparting a bias toward unity in the coefficient of any estimated relationship.

Thus I can imagine running regressions like Feldstein and Horioka have done, using, say, observations on state-level saving and investment rates, and getting results that are very similar to theirs. But this would not mean that the states were not part of a single and well-functioning capital market, or that their economies should be treated as at least quasi closed for the purposes of studying tax incidence and other things.

Now the same line of reasoning as was used to hypothesize that the correlation between saving and investment rates would be higher for states than for cities, and higher for cities than for blocks, can be used to conjecture about countries. I have chosen here to focus on a particular variable—the gross capital flow out of or into a country, expressed as a fraction of its gross investment. This variable can be easily calculated from data presented in Table 5 of the World Bank's *World Development Report, 1979.* It reports gross domestic saving and gross domestic investment rates for over 100 countries, for 1960 and 1977. The variable I calculated, x_{it}, is defined as the difference between these two, expressed as a fraction of the gross investment rate. My conjectures are

1) That this variable will have higher variability for small and poor countries than for larger and more wealthy ones;

2) That the absolute value $|x_{it}|$ will tend on the average to be higher for small and poor countries than for larger and more wealthy ones.

Table 2 compares the *OECD* countries with middle- and low-income countries. The pattern is very clearly one of low values of x_{it} and relatively little variance for the rich countries, and of increasing absolute value and increasing variance as one moves to the poorer countries. Poor countries are more like city blocks, rich countries more like states! The point to be borne in mind here is that the evidence of the Feldstein-Horioka paper was assembled from the *OECD* countries only. Casting the net wider would have surely thrown up indications of much greater divergence between saving and investment rates. (I believe that a net investment, net saving comparison would be more meaningful, since it is perfectly "nor-

TABLE 2—DISTRIBUTION OF x_{it} (=(GROSS DOMESTIC SAVINGS/GROSS DOMESTIC INVESTMENT) − 1), BY COUNTRY CLASS

	First Quartile	Median	Third Quartile	Interquartile Range	Median Absolute Value $\|x_{it}\|$
1960					
OECD Countries	−.085	−.042	+.028	.113	.07
Middle-Income Countries	−.31	−.116	.09	.40	.23
Low-Income Countries	−.53	−.20	0	.53	.32
1977					
OECD Countries	−.109	−.070	0	.109	.09
Middle-Income Countries	−.25	−.15	−.04	.21	.19
Low-Income Countries	−.58	−.34	−.10	.48	.42

TABLE 3—MEDIAN VALUES OF $|x_{it}|$ BY COUNTRY CLASS AND POPULATION SIZE

	Population 0–5		Size 5–20		(millions) 20+
1960					
Low-Income Countries	.40	>	.36	>	.14
	∨		∨		∨
Middle-Income Countries	.31	>	.28	>	.12
	∨		∨		∨
Industrialized Countries	.08	>	.04	<	.06
1977					
Low-Income Countries	.71	>	.36	>	.23
	∨		∨		∨
Middle-Income Countries	.20	>	.19	<	.20
	‖		∨		∨
Industrialized Countries	.20	>	.10	>	.05

Note: Results different from conjecture have been circled.

mal" for most types of capital to be replaced as they wear out. While I have not been able to make the necessary adjustments, it is a good guess that both the values of the x_{it}, and the measures of their variability, would range between 1.5 and 2 times the figures appearing in Table 2. Thus a typical middle-income country might normally be using external financing for something like a third of its net investment, and for a typical poor country this figure might well be more than half.)

Table 3 elaborates on the above picture by breaking down countries by population size as well as by per capita income level. Here the variable reported is the median $|x_{it}|$. This choice permits divergences between saving and investment behavior to be reflected, even for country groups whose x_{it} are symmetrically distributed around zero. Indeed the median $|x_{it}|$ is in a certain sense a measure of variability, being analogous to an absolute moment around zero.

The results in Table 3 are quite in line with the conjectures presented earlier. The inequalities that conform to our conjectures are not circled; those at variance with the conjectures are circled. Of twelve possible inequalities in each panel of the table only one (in each panel) shows the wrong sign, and one (for 1977) shows an equality where an inequality was expected. Moreover the two cases of wrong sign exhibit very little difference in magnitude. So I am content to consider that the results are broadly compatible with the conjectures. This leads me

to conclude that the World Bank data on saving and investment rates are broadly compatible with the notion of a reasonably well-functioning world capital market.

III. An Attempt at Interpretation

The evidence presented above tends to bolster the "capital market" as against the "quasi-closed economy" interpretation, but I must confess that I am not particularly at ease with all the implications of a really well-functioning world capital market. In particular, my own intuition does not want to accept the notion that increments of investment activity are in all or nearly all countries effectively 100 percent "financed" by funds flowing in from abroad, and that increments in saving simply spill out into the world capital markets.

I find the analogy to a hydraulic system with perhaps a viscous fluid, in which the pipes are partially clogged, and in which some vessels are separated by semipermeable membranes, to be more consonant with my image of the world than the alternative analogy to a hydraulic system where the water flows freely through the system and, essentially instantaneously, finds the same level everywhere.

This is enough to permit shifts in investment functions to influence domestic savings rates, and shifts in savings functions to influence domestic investment rates over some period of time. It is probably enough to rationalize the time-series results that Feldstein and Harioka obtained. But one is struck with the implication that in the long run even a viscous fluid will find its level. So the world market would still be the marginal source of investible funds and the marginal dumping ground for excess saving—in the long run if not in the short.

It is not easy to escape from that implication. The best I can do is conjure up a reason why the supply curve of funds facing a country would be upward sloping. The argument goes like this: If lenders and borrowers perceived default risk in exactly the same way, there would be a presumption that the risk-free rate would govern everybody's behavior. If a consol paying \$10 a year would sell for \$100 if risk free, then a consol paying \$20 on December 31 of each year if a flipped coin turns up heads, and zero if it turns up tails, should also sell for \$100. The event of "default" is perceived as a benefit by the borrower, and as a cost by the lender; it enters symmetrically in their perceptions. They see the interest rate as 10 percent in both cases.

But if the borrower is not so callous as to perceive default as a benefit (and indeed it surely has its genuine pains in most cases), or if he places a lower probability on default than does the lender, this can put the borrower in a quasi-monopsony position vis-à-vis the capital market. The true supply curve of risk-free funds facing any one borrower is a horizontal line at r_p, the risk-free rate. But the supply price of funds (F) is an upward-rising function, $r_p+\delta_L(F)$, where δ_L is the risk premium perceived by the lender(s). Now if the borrowers' perception of risk was the identical function $\delta_L(F)$, and if he treated nonpayment as a benefit to himself, then he would act in terms of the risk-free rate r_p, as in the example with the consols. But if the borrower perceives this "benefit-from-default" function as $\delta_B(F)$, then his true cost of credits is $r_p+[\delta_L(F)-\delta_B(F)]=r_p+\gamma(F)$. His marginal cost of credit would presumably be the curve marginal to the $r_p+\gamma(F)$ function.

Does this point the way out of our dilemma? It's a slim reed, perhaps, but I find myself less disturbed by it than by either the extreme capital market or the extreme quasi-closed economy case. It rings true in terms of what banks and others will lend to you and me. It rings true in terms of the way the interest rates that business firms pay will vary with their debt-equity ratios. And I believe it rings true in terms of the premia for "country risk" that the world capital market assigns, at least to many *LDC*s.

But these are, in a sense, just rationalizations. My guess is that most of us feel, from what we have observed of the world economy, that capital just doesn't flow as freely internationally as the capital market interpretation would have it. What I have just presented is one way of rationalizing a sys-

tem where increments in investment are partially financed from the home market, and partially from the international market, and where increments of saving find their way to finance some new investments at home, and some abroad. My question is, is there any alternative plausible way of rationalizing such a system?

REFERENCES

M. Feldstein and C. Horioka, "Domestic Savings and International Capital Flows," *Econ. J.*, forthcoming.

A. C. Harberger, "Perspectives on Capital and Technology in Less Developed Countries," in M. J. Artis and A. R. Nobay, eds., *Contemporary Economic Analysis*, London 1978.

______, (1972a) "On Estimating the Rate of Return to Capital in Columbia," in his *Project Evaluation*, Chicago 1972.

______, (1972b) "Investment in Men versus Investment in Machines: The Case of India," in his *Project Evaluation*, Chicago 1972.

G. Jenkins, "Analysis of Rates of Return from Capital in Canada," unpublished doctoral dissertation, Univ. Chicago, Dec. 1972.

D. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Surv. Curr. Bus.*, May 1972, *52*, 31–64.

C. G. Langoni, "A Study in Economic Growth: The Brazilian Case," unpublished doctoral dissertation, Univ. Chicago, Dec. 1970.

R. de F. Novaes, "Foreign Investments in Brazil—An Economic Analysis," unpublished doctoral dissertation, Univ. Chicago, June 1974.

A. H. Petrei, "Rates of Return to Physical Capital in Manufacturing Industries in Argentina," unpublished doctoral dissertation, Univ. Chicago, Sept. 1971.

George J. Stigler, *Capital and Rates of Return in Manufacturing Industries*, Princeton 1963.

World Bank, *World Development Report, 1979*, Washington 1979.

DISCUSSION

BEN S. BERNANKE AND JEREMY I. BULOW, Stanford Graduate School of Business: The first order of business is to commend Barbara Fraumeni and Dale Jorgenson for their massive but high-quality data-gathering effort, of which the paper today is only a small manifestation. Their estimates of industry capital stocks and returns will be useful to researchers in several areas.

We were curious to see how the nominal return estimates presented by Fraumeni-Jorgenson (hereafter F-J) would compare with the implicit estimates of investors who participated in the stock market during the same period (1948–76). Our source was Standard and Poor's (*S-P*) *Market Analysts' Handbook.* We found there industries or sets of industries coinciding with twenty-one of the SIC groupings used by F-J. (See our Tabulation, which is a comparison of *S-P* and F-J estimates of nominal industry rates of return for 1948–76. The * denotes 1957–76 data.) For each *S-P* industry we calculated the average dividend yield and the average growth of the stock price index over the period. The sum of these two numbers was taken as an approximation to the actual nominal return experienced in the market. (Note that a really careful analysis would take into account variations in dividend yield over time, interindustry differences in debt-equity ratios, and the return to intangible stocks corresponding to flows like advertising or research and development.) Where several *S-P* industries corresponded to one SIC group, we aggregated using relative market values as weights.

	Industry	*S-P*	F-J
1.	Coal	.146	.142
2.	Crude petroleum	.101	.124
3.	Food	.101	.103
4.	Tobacco	.115	.135
5.	Textiles	.087	.090
6.	Apparel	.064	.105
7.	Paper	.128	.128
8.	Printing and Publishing	.101	.107
9.	Chemicals	.114	.132
10.	Rubber	.126	.105
11.	Primary metal	.106	.090
12.	Fabricated metal	.123	.105
13.	Elec. machinery	.137	.121
14.	Motor vehicles	.131	.285
15.	Railroads	.076*	.076*
16.	Air transport	.068	.020
17.	Tel. & Tel.	.088*	.178*
18.	Radio-TV	.154	.151
19.	Elec. utilities	.088	.131
20.	Gas utilities	.106	.145
21.	Retail trade	.101	.102

Basically, we found a good match between the *S-P* and F-J results. The rank correlation is .57, rising to .79 when the three regulated utilities are left out. Without the utilities, this is higher than the correlation of the *S-P* ranking and an ordering based on standard accounting rates of return (that correlation was .59). That is, F-J predict the ordering of stock market returns over industries better than does information on earnings and book values. F-J and *S-P* also agree reasonably well in absolute terms: eight of the twenty-one industries match within .5 percent or so, and seven more agree within 2.5 percent.

This leaves six industries where the *S-P* and F-J results differ by more than 2.5 percent. These are of interest. Three of the exceptions are the utilities—gas, electric, and telephone. Here, F-J estimate a higher return than *S-P* in each case. F-J's revaluation procedure is probably inappropriate for these industries, where actual returns are regulated and are related to historical costs. That is our justification for excluding the utilities from our rank correlation calculation.

Two of the other three exceptions are the F-J outliers—air transport and motor vehicles. While the market agrees that these are low- and high-profit industries, respectively, it places them much closer to the mean in return than do F-J. Indeed, in general the *S-P* results cut off the tails of the F-J return distribution: the *S-P* range is from 6.5 to 15.5 percent compared to F-J's 2 to 28.5 percent. This weakens the F-J claim that there are wide variations in return across industries. Of course, this exercise does not tell us whether F-J or the market has perceived the facts more accurately.

Besides presenting these numbers, F-J offer several conclusions. One of these is that the United States could have had a much higher return over the period by allocating capital to motor vehicles, say, rather than to air transport. This is an arithmetic fact, given their estimates. But do they mean this as a simple statement of hindsight, or are they suggesting that there were exploitable differentials *ex ante*? One may argue against the latter interpretation on several grounds: 1) An extensive empirical literature suggests that *U.S.* capital markets are efficient; 2) As the stock market data show, the variance of *ex post* returns may be smaller than F-J estimate; 3) Risk differentials across industries make some dispersal of returns desirable in any event.

Improving Productivity Measurement

By ALBERT REES*

Productivity is measured with error, and this error can at times be substantial. Moreover, some of the error can be reduced at costs that are reasonable given the importance of productivity change to our economy.

The most widely used broad measure of productivity change, the *BLS* measure of output per hour of labor in the private business sector, encounters sources of error on both the output and the input sides. For output, two sources of error are particularly important. The first is that for some parts of the private business sector, comprising about 5 percent of the total, real output is not measured independently and input proxies are used instead. This occurs, for example, when there is no suitable price series available for deflating the current dollar value of output, and average hourly earnings of workers in the industry are used as the deflator. Such proxies are used in parts of the service industries and in some types of construction. The use of input proxies biases the productivity measure toward showing no change. This source of error is one of the most difficult to overcome and the possible methods of dealing with it are not at all obvious.

It should also be noted that the lack of an independent measure of real output is what prevents the inclusion of general government and nonprofit institutions in the scope of the broadest productivity measures. The exclusions are most unfortunate, since the excluded sectors are very large. More research is needed on appropriate ways of measuring final output and productivity in government and such organizations as universities and nonprofit hospitals, but the task will not be an easy one.

The second source of error on the output side of our broad productivity measures lies in the inadequacy of some of the price deflators used to convert measures of output valued in current dollars into measures of real output. These deflators omit some very important classes of products, such as computers and aircraft, in which technological change has been extremely rapid. To use as proxies the prices of other classes of more conventional products, where technological change has been far less rapid, means that we systematically overstate the rise in prices and correspondingly understate the growth of output and productivity. It is inappropriate and perhaps not feasible to use ordinary specification pricing of the kind used for common commodities to price such rapidly changing products as computers. However, several economists outside government have used hedonic price indexes for this purpose. There is no objection in principle to using hedonic price indexes as components of the official price deflators; it is already done for single family houses.

A similar understatement of the growth of output and productivity occurs when certain types of quality improvement take place in products that are included in the price indexes. At present, official price indexes are adjusted appropriately for quality improvements that require increased production cost, such as the inclusion as standard equipment in an automobile of equipment that was previously optional at extra cost. The portion of any associated price increase that represents the cost of producing the equipment in question is treated as a quality improvement rather than a price increase.

*The Alfred P. Sloan Foundation. My views reflect my recent experience as chairman of the Panel to Review Productivity Statistics of the National Academy of Sciences; however, this paper does not necessarily reflect the position of the Panel.

Similar adjustment is *not* made, however, when improved quality results from technological change that does not add to production cost (as distinguished from the cost of research and development). Methods of dealing with this kind of quality change must be more complex than those used to deal with adding radios or automatic transmissions to cars. However, the problem is by no means an insoluble one, and more resources should be devoted to it.

Let us turn next to the input side of the labor productivity measure. Labor input can be divided into its two principal components, the number of workers and average hours per worker. The first of these does not seem to present serious measurement problems, but hours per worker does. For employees, who make up the great bulk of the work force, we now measure the hours paid for by employers rather than the hours spent at the workplace. The difference between the two hours concepts consists of paid leave, which is largely vacations, holidays, and paid sick leave. Since most employees are not engaged in production when they are not at the workplace, hours paid for is not the appropriate concept for productivity measurement. The inappropriateness of the hours concept used would not bias productivity measures if paid leave were a constant fraction of paid hours, but it is not. Since World War II the ratio of paid leave to total hours paid for has grown substantially.

A *BLS* task force has proposed that an annual survey of hours spent at the workplace be administered to a subsample of the employers who now report hours paid. The ratio of hours worked to hours paid obtained from the subsample could then be used to adjust the data on hours paid from the main sample. This is a sound and workable proposal and deserves support.

Some recent research by Frank Stafford and Greg Duncan reports an increase in the time spent at the workplace that is not used in doing work—that is, an increase in the consumption of leisure on the job. This could be another source of downward bias in measured productivity, but it is a much harder one to measure routinely than paid leave time.

The second major difficulty with the hours data now used is that they cover only production workers in manufacturing and mining and only nonsupervisory workers elsewhere. The hours of the excluded groups, nonproduction and supervisory workers, are estimated by various rules of thumb. These groups constitute 18 percent of the workers on all private nonagricultural payrolls and 28 percent of workers in manufacturing and mining. The collection of annual data on the hours of nonproduction and supervisory workers would greatly reduce reliance on crude methods of estimation.

The *BLS* measures of productivity all use unweighted measures of labor input, in which an hour of work done by an unskilled laborer is treated as equivalent to an hour of work by a craftsman or a professional worker. Many economists outside government use measures of weighted labor input, in which different kinds of labor are weighted by measures of hourly earnings to reflect differences in the marginal productivities of different classes of workers. Changes in such measures reflect gains in efficiency resulting from the replacement of skilled by less skilled labor.

Not all differences in earnings reflect differences in skill—some may reflect discrimination in labor markets or other imperfections in these markets. For this reason there is not agreement among economists on what kinds of categories to use in assigning weights. It would nevertheless be desirable to develop some official measures of weighted labor inputs to help us better understand the sources of change in the unweighted measures.

The productivity measures produced by the federal government, except those for agriculture, are all measures of labor productivity. In contrast, economists outside government have for many years made estimates of total factor productivity, or, as the Panel to Review Productivity Statistics called it, multifactor productivity. Such estimates include inputs of capital services

in the denominator. For particular sectors of the economy, they also include the inputs of intermediate goods and services purchased from other sectors. As is well known, changes in measures of multifactor productivity reflect changes in efficiency in the use of all measured inputs rather than in the use of labor inputs alone. A measure of labor productivity can sometimes rise only because capital services or intermediate goods are being substituted for labor.

It is time for the federal government to begin to produce official measures of multifactor productivity consistent with its major measures of labor productivity. A comparison of the two kinds of measures will then enable us to see easily what part of the change in measured labor productivity is due to factor substitution.

To improve the accuracy of measures of multifactor productivity we need better measures of capital stock, including inventories, and better measures of inputs of intermediate goods. Measures of inputs of intermediate services, now entirely lacking, are also needed. Detailed recommendations for the collection of such data in the quinquennial economic censuses were made recently by the Gross National Product Data Improvement Project.

There is now either little information in existence on productivity at the plant or company level in the United States or little of what information exists is in the public domain. The fragmentary evidence available suggests that there are very large differences in the productivity of establishments within the same 4-digit industry. It would seem desirable both to increase efforts to measure productivity at the establishment level and to increase access to the measures now regarded as proprietary, since measurement at the establishment level is more likely than broader measures to assist and direct efforts to improve productive efficiency.

In closing, I turn to the relationship between errors in productivity measurement and the recent slowdown in measured productivity growth. The major sources of error in productivity measurement clearly seem on balance to bias the measures downward —that is, better measurement would produce time-series measures of productivity that rise more over long periods than the ones we now use. In particular, downward bias is produced by the errors caused by the use of input proxies to measure output, the failure to capture all kinds of quality improvements in price deflators, and the use of hours paid instead of hours spent at the workplace in the labor input measure. The presence of errors that bias a measure downward, however, is not sufficient to cause a decrease in the measure's rate of growth. To cause such a decrease, a downward-biasing error would have to be getting larger through time. I know of nothing that suggests that this is true of the errors I have discussed, and a good deal that suggests the opposite. Ten years ago we used more input proxies to measure output, and had poorer adjustments for quality change than we do now. Moreover, the growth of paid leave was probably more rapid before the mid-1960's than it has been since. Thus a better measure of productivity change would have risen more than the present ones over the whole period since World War II; but the deceleration in its rate of growth since the mid-1960's would probably have been even more pronounced than that in the existing measures. In short, the lag in measured productivity growth seems to me to be entirely a real phenomenon and in no part attributable to measurement error. Indeed, if anything the deceleration has been understated.

$R\&D$ and the Productivity Slowdown

By Zvi Griliches*

The question I shall address in this paper is: Can the slowdown in productivity growth be explained, wholly or in part, by the recent slowdown in the growth of real $R\&D$ expenditures? But first we have to review the following questions: 1) What is to be explained? Which productivity and what slowdown? 2) What is the mechanism by which $R\&D$ could have contributed to this slowdown? 3) What did happen to $R\&D$ in the relevant period? Besides traversing this somewhat familiar ground and reviewing some of the recent literature on this topic, I shall also report on some estimates of my own.

The direct answer to the opening question is "probably not." But how we get there needs documenting and may prove instructive on its own merits.

I

There are several productivity "slowdowns" which may be candidates for an explanation. The literature defines two slowdowns in the growth of labor productivity: 1965–73 and 1973–78. The first slowdown occurs almost entirely outside of manufacturing. The second is pervasive and steep but seems to be associated with the aftermath of the energy crisis which began in 1974 and the deep recession of 1974–75. If one looks at total factor productivity the picture is murkier. In Frank Gollop and Dale Jorgenson there is no evidence that total factor productivity grew more slowly in manufacturing in 1966–73 than earlier (1960–66). Among the twenty 2-digit industries reported by them eleven had higher (or equal) rates of growth in the later period and nine had lower ones. John Kendrick estimates that the rate of growth of total factor productivity in manufacturing in 1966–73 was below that in 1957–66 but comparable to the average rate of growth between 1948 and 1957 (2.1, 3.0 and 2.0, respectively). Looking at ninety-five *BLS* growth sectors in manufacturing (most of which are at the 3-digit SIC level), which are the numbers I have been analyzing recently, one cannot discern a clear slowdown in labor productivity before the mid-1970's.

In what follows I shall concentrate primarily on manufacturing because this is where one would expect to be able to observe best the effects of $R\&D$ on productivity. In most other areas, such as services or government, output is not measured distinctly enough from input to be worth analyzing in any detail. The only other major sectors where measured productivity could be affected by $R\&D$ expenditures in the longer run are agriculture, communications, transportation, and public utilities. There has been no recent productivity decline to speak of in the first three, while the decline in the utilities sector is largely due to the reduction in capacity utilization associated with the energy crisis.[1]

II

Research and development is an investment flow. What affects output is presumably some cumulated stock of the previous results of such investments. Since such results are not easily measurable, most growth accountants have constructed some stock of $R\&D$ capital measure: $K_t = \Sigma w_i R_{t-i}$, where the "stock" K is a function

*National Bureau of Economic Research and Harvard University. This research has been supported by NSF grants No. SOC78-04279 and PRA79-13740. I am indebted to Alan Siu for very able research assistance and to the Bureau of Labor Statistics, Office of Economic Growth, for making its unpublished data series available to me.

[1]Because of space limitation I am not documenting the various statements made in the text. They are consistent with the facts as reported by Edward Denison, Gollop and Jorgenson, Kendrick, and J. R. Norsworthy, M. J. Harper, and Kent Kunze.

TABLE 1—AVERAGE GROWTH RATES WITHIN MANUFACTURING INDUSTRIES
(Percent per Year)

	Output per Man-Hour	R&D Stock $\delta=0$	R&D Stock $\delta=.1$	R&D Stock $\delta=.2$
1960–65	3.4	10.3	8.5	7.0
1965–73	2.8	7.9	6.4	5.6
1973–77	1.4	5.6	3.0	1.4
	Deceleration			
1960–65 to 1965–73	0.6	2.4	2.1	1.4
1965–73 to 1973–77	1.4	2.3	3.4	4.2

of past investments R, and the ws reflect the assumed lag and depreciation schemes. Given such a measure, the contribution of K to output growth is measured by γk, where γ is the elasticity of aggregate output with respect to $R\&D$ capital and k is its rate of growth. An alternative form is given by $\rho NR/Q$, where ρ is the gross rate of return to $R\&D$ investments, NR is the *net* investment in research capital and Q is total or sectoral output. Thus before we can proceed to an estimate of the effects of $R\&D$ on growth we need: 1) an estimate of γ or ρ; 2) a measurement of the *relevant* rate of $R\&D$ expenditures; and 3) evidence or an assumption about its lag structure and depreciation pattern.[2]

The $R\&D$ to GNP ratio peaked at about 2.9 percent in 1964 and declined slowly to about 2.3 percent in 1975. Not all of this $R\&D$ is contributing to productivity growth as it is currently measured. Much of it is spent on defense and space exploration, on health and environment, and on goods and services (such as computers) where quality improvements brought about by such expenditures are not captured in our national accounts. Since much of the slowdown in $R\&D$ occurred in these sectors, the slowdown in $R\&D$ that could have had a measurable impact was not as large as the crude figures might indicate. I have previously estimated (1973) that only about half of total $R\&D$ as measured is likely to affect measured productivity and that only about half of the remainder represents net additions to the stock of research results. If these ratios have remained constant, the "effective" $R\&D$ to GNP ratio may have declined from about 0.72 to .57, or about .25 percentage points. Even assuming a relatively high rate of return of 40 percent would account at most for a .1 percent decline in the rate of growth.

But how about its timing? The fact that $R\&D$ investments in constant dollars peaked in the mid-1960's does not imply that the associated stock measures peaked at the same time. Table 1 reports average growth rates in several measures of the stock of applied research and development capital and the associated rate of growth of output per man-hour in manufacturing (adjusted for interindustry shifts). It does indicate a clear decline in the rate of growth of $R\&D$ capital of about 3 to 6 percentage points, depending on the exact comparison period and the particular series chosen for the comparison.

III

To assess the impact of such a substantial decline in the rate of growth of $R\&D$ capital on productivity growth we need an estimate of the elasticity of aggregate or sectoral output with respect to changes in the $R\&D$ capital. In previous work (1979, 1980) using 1957–65 data on 883 large *U.S.* corporations, I estimated this elasticity to be about .06 (corresponding to a no-depreciation concept of $R\&D$ capital). This estimate together

[2]See my 1979 paper for a more detailed exposition. See also my 1973 paper, Edwin Mansfield, and Nestor Terleckyj for earlier applications of this kind of framework.

with a lowering of the *R&D* capital rate of growth by about 2.3 percent (see Table 1) imply a contribution of about .14 percent to the productivity slowdown in manufacturing, accounting for about one-tenth of it.

This could be an underestimate for two reasons: (a) my earlier estimates are based on firm data and hence do not capture social returns and the spillover effects of *R&D*; and, (b) they are based on an earlier period, when everything was growing together. In the more recent period there has been more variance in the *R&D* variable and perhaps one could get better and also possibly higher estimates if the analysis were extended to the 1970's.

I turn, therefore, to an analysis of *R&D* and productivity at an approximately 3-digit SIC level, using the newly released *BLS* growth sector output, man-hours, and capital stock data (see *BLS*, 1979a,b), and the *NSF* data on Applied Research and Development (*AR&D*) expenditures by product fields. The ninety-five *BLS* sectors were aggregated into thirty-nine sectors, and the twenty-nine *NSF* product fields were disaggregated to match. The *R&D* series were deflated and cumulated using a declining balance depreciation scheme.[3] A capital service flow measure was defined as Depreciation in constant dollars plus .08×Net Stock of Fixed Capital. The *BLS* numbers are for gross output, not value-added, but I shall proceed on the assumption that the two moved proportionately for most of the period in question. Both the numbers and the measures constructed from them are rather crude. The purpose, however, was not to do a detailed industry-level total factor productivity analysis, but rather to see whether there is some prima facie evidence for *R&D* being the major culprit in the recent slowdown.

The basic approach is to allow for industrial differences in both the average level of productivity and in the capital elasticity. This is accomplished by estimating everything "within" industries, that is, after subtracting the mean levels of each variable within each industry, and by multiplying the capital variable in each industry by its estimated share in value-added. Table 2 presents the major results of this analysis using the no-depreciation version of *R&D* capital.[4] If one starts with a definition of total factor productivity (*TFP*) and allows only *R&D* to affect it (besides time dummies), the estimated coefficients are high and very significant. If one adds, however, other variables to the equation, such as the average age of capital and total man-hours (reflecting the short-run phenomenon of increasing returns to scale during the business cycle), the *R&D* coefficients decline, as does their statistical significance. If we do not impose the *TFP* framework and allow the capital variable to have its own coefficient, the results are better both in terms of fit and in the significance of the *R&D* variables, though the capital variable gets a coefficient of less than half its expected size. When the period is broken into two, 1959–68 and 1969–77, the results for the first period are strengthened, with the estimated coefficient being about the same (.07) as I had found earlier at the micro level, but there is nothing to be found in the second period. The contribution of *R&D* goes to zero or at least cannot be discerned in these data and in this period.

In part this could be a data problem: price indexes get differentially bad as inflation begins to accelerate, the proportionality

[3]The *AR&D* data by product field were used rather than the *R&D* industry figures because (a) they are more compatible with the establishment-based productivity series and (b) they provide somewhat more industrial detail. Basic research is not broken down by product field and hence is not included in these stock measures. Since it accounts for only 3 percent of total *R&D* in industry, its inclusion would not affect the overall picture. Note that one could quarrel with a measure of social *R&D* capital that may decline just due to the passage of time. But to open this issue here would take us too far afield.

[4]Experiments with (geometric) depreciation rates of 10, 20, and 30 percent per year led largely to the same results. The data can hardly distinguish between them, showing a slight preference for 0 or 10 percent. The only noticeable difference in the results are negative though insignificant *R&D* coefficients for the 1964–77 period when higher depreciation rates are used.

TABLE 2—$R\&D$ STOCK COEFFICIENTS IN VARIOUS SPECIFICATIONS: WITHIN THIRTY-NINE $U.S.$ MANUFACTURING SECTORS
($N=741$)

Period and Dependent Variable	$R\&D$ Stock Coefficient[a]	$S.E.E.$	Other Variables in Equation
TFP, 1959–77	.058 (.030)	.124	None
TFP, 1959–77	.029 (.027)	.110	Age C, $MHRS$
$LPROD$, 1959–77	.044 (.026)	.106	$(1-LS)\cdot L(SFC/MHRS)$, Age C, $MHRS$
$LPROD$, 1959–68	.067 (.029)	.064	$(1-LS)\cdot L(SFC/MHRS)$, Age C, $MHRS$
$LPROD$, 1969–77	.026 (.046)	.078	$(1-LS)\cdot L(SFC/MHRS)$, Age C, $MHRS$

Notes: $TFP=LPROD-(1-LS)\ [LSFC-MHRS]$; $LPROD=$ *Log* output per man-hour; $(1-LS)\ [L(SFC/MHRS)]=(1-$Average 1971–75 labor share in value-added) (*Log* service flow measure of fixed capital minus *log* total man-hours); $R\&D$ Stock: *log* (cumulated sum of deflated applied research and development expenditures from 1959 on plus 1958 initial value). 1958 initial value$=R_{58}/g$, where g is the estimated rate of growth of R in a particular industry during 1959–64; $S.E.E.$ =standard deviation of residuals. All equations contain an intercept and year dummies. "Within" means that all variables are measured around their respective industry means.

[a]Standard errors are in parentheses.

between gross output and value-added begins to break down in this period, and there are rather large swings in capacity utilization unaccounted for by our theories and by the measures used here. I suspect, however, that the effect may be real. First, various attempts to improve upon these results by allowing for differential materials and energy intensity of the different industries, adding capacity utilization measures in the 1967–77 period, and by throwing out industries whose price deflators are of dubious quality, led to no appreciable improvement. Second, the lack of published results on the contribution of $R\&D$ to productivity based on post-1968 data leads me to suspect that this is not an isolated finding or peculiar to the specific data set used here. I know of only three studies whose results are based on data which go past 1968: M. Ishaq Nadiri, Roger Brinner, and Carson Agnew and Donald Wise. All three studies include parts of the later period in their total sample but only Nadiri examines it separately. He must resort, however, to ridge regression techniques to obtain sensible-looking results and does not include a separate trend variable, attributing all of TFP growth to $R\&D$. Brinner's estimates are based on total domestic GNP, including many sectors where productivity is either not measured at all (such as residential capital formation) or measured very badly (such as services). His estimates are of similar order of magnitude (.06) but are not very robust to the inclusion of other variables (they fall to .03 when labor is adjusted for quality change). The Agnew-Wise study fits TFP estimates for eleven 2-digit $U.S.$ manufacturing industries for the period 1957–75 to various $R\&D$ measures and $R\&D$ spillover measures and gets essentially nothing.

If these findings are to be taken at their face value, they imply a larger effect of $R\&D$ on the slowdown with the effect coming not so much from the slowdown in $R\&D$ as from the collapse in the productivity of $R\&D$. If one assumes that $R\&D$ capital was previously growing at a rate of 6 percent per year and the associated $R\&D$ output elasticity was .07, then its contribution to the rate of growth of productivity in manufacturing

was .42 percent per year in the earlier period. The total disappearance of this contribution could account for more than a quarter of the recent productivity slowdown.

How can one explain this collapse of the *R&D* coefficient in the 1970's? It is possible that a large fraction of recent *R&D* investments has been diverted to finding ways of complying with various new environmental and other regulatory constraints. But that should have not have happened across the board. Not all industries have been subject to the same regulatory pressures. Nor should it depreciate all of the "older" *R&D* investments down to zero. It is also possible that much of the effect of past *R&D* is embodied in new equipment, and a slowdown in capital growth may also induce a decline (a postponement) in the effect of *R&D* on productivity. This may be testable and would imply that a pickup in investment would also induce a recovery in the *R&D* coefficient. The most likely explanation is one of confusion: the large energy price shocks, the resulting fluctuations in capacity utilization, the substantial increase in uncertainty about future absolute and relative prices may have forced many firms away from their long-run production frontiers. What we see in the data are not movements along the technological frontier, and hence they should not and cannot be attributed to a variable whose role is to shift this frontier outward.

The other point to remember is that even though the measured effects of *R&D* on measured productivity may be small, its true effects may be quite a bit larger. First, we have yet to learn how to measure the spillover effects of *R&D* within and across industries.[5] Second, much of past and current *R&D* is spent on socially valuable activities such as our health and the health of our environment, items that are not valued positively in the national accounts as currently constituted. Finally, *R&D* is a chancy and fickle process. Even if it has run into a dry spell, this does not imply that current expenditures may not have future returns or that there are no major productivity gains already on the drawing boards. All substantive surveys of new technologies and new technological possibilities seem to contradict the notion that we have exhausted our innovation possibilities. Thus I interpret my lack of findings as reflecting data difficulties and the turmoil of the times rather than a true underlying trend shift. In any case, it is unlikely that the recent productivity slowdown can be blamed primarily on the *R&D* slowdown. If anything, causality may run in the other direction.

[5]See my 1979 paper for more discussion, also Agnew-Wise and M. A. Schankerman for examples of such attempts. (There has been no convincing showing of spillover effects using post-1968 data.)

REFERENCES

C. E. Agnew and D. E. Wise, "The Impact of *R&D* on Productivity: A Preliminary Report," paper presented at the Southern Economic Association Meeting 1978.

R. Brinner, "Technology, Labor, and Economic Potential," study no. 29, Data Resources, Inc. 1979.

E. F. Denison, "Exploration of Declining Productivity Growth," *Surv. Curr. Bus.*, No. 8, 1979, *59*, Part II, 1–24.

F. M. Gollop and D. W. Jorgenson, "U.S. Productivity Growth by Industry, 1947–73," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, 1980, 17–124.

______ **and** ______, "U.S. Economic Growth 1948–1973," unpublished manuscript, Univ. Wisconsin-Madison 1979.

Z. Griliches, "Research Expenditures and Growth Accounting," in B. R. Williams, ed., *Science and Technology in Economic Growth*, London 1973, 59–95.

______, "Returns to Research and Development Expenditures in the Private Sector," in John W. Kendrick and Beatrice Vaccara, eds., *New Developments in Productivity Measurement*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 44, 1980, 419–54.

______, "Issues in Assessing the Contribution of Research and Development to

Productivity Growth," *Bell J. Econ.*, Spring 1979, *10*, 92–116.

J. W. Kendrick, "Productivity Trends and the Recent Slowdown," in William Fellner, ed., *Contemporary Economic Problems*, Washington 1979, 17–69.

R. E. Kutscher, J. A. Mark, and J. R. Norsworthy, "The Productivity Slowdown and the Outlook to 1985," *Mon. Labor Rev.*, May 1977, *100*, 3–8.

E. Mansfield, "Rates of Return from Industrial Research and Development," *Amer. Econ. Rev. Proc.*, May 1965, *55*, 310–22.

M. I. Nadiri, "Contributions and Determinants of *R&D* Expenditures in the U.S. Manufacturing Industries," Nat. Bur. Econ. Res. work. paper no. 360, New York 1979.

J. R. Norsworthy, M. J. Harper, and K. Kunze, "The Slowdown in Productivity Growth: Analysis of Some Contributing Factors," *Brookings Papers*, forthcoming.

M. A. Schankerman, "Essays in The Economics of Technical Change," unpublished doctoral dissertation, Harvard Univ. 1979.

Nestor E. Terleckyj, *Effects of R&D on the Productivity Growth of Industries: An Exploratory Study*, Washington 1974.

Bureau of Labor Statistics, (1979a) *Time Series Data for Input-Output Industries*, Bull. 2018, Washington 1979.

———, (1979b) *Capital Stock Estimates for Input-Output Industries: Methods and Data*, Bull. 2034, Washington 1979.

Sectoral Productivity Slowdown

By M. ISHAQ NADIRI*

The recent dramatic slowdown of *U.S.* aggregate productivity growth has been the subject of intensive discussion in the literature. One source of the slowdown may be the substantial changes in the industrial composition of output, employment, capital accumulation, and resource utilization. It is fairly evident that the slowdown is widespread throughout all sectors of the economy, but what is not quite obvious are the factors which have been responsible for this pervasive phenomenon. Some available studies attribute the retardation of the aggregate productivity growth to sectoral shifts in the composition of output and employment.[1] However, the intersectoral shift in employment can only explain a fraction of the aggregate productivity slowdown. Also, what is not readily known is the explanation for the changes in growth of the sectoral productivity.

In this paper, I shall concentrate on two questions: What set of factors explains, at least partially, the growth of labor productivity in different sectors during the postwar period? Is the weakening of the same set of factors accountable for the slowdown of the sectoral and aggregate productivity growth since 1973? What follows describes very briefly a model of labor productivity growth. The model is estimated using sectoral and aggregate data for the period 1949–78. The data covers output, employment, capital stock, stock of *R&D*, level and rate of change of utilization rate, and a proxy for disembodied technical change.

I conclude from my analysis that the growth rate of the capital-labor ratio, the utilization rate and its rate of change, and the growth stock of total *R&D*, go a fair distance in explaining the pattern of sectoral and aggregate productivity growth. The weakening of the same set of forces is shown to have contributed to the recent slowdown in these growth rates.

I

Before describing our model, it is useful to note briefly that during the postwar period, substantial changes have taken place in patterns of sectoral growth of output, employment, capital stock, and *R&D* accumulation, and degree of resource utilization. A brief look at the growth rates over the subperiods, 1948–55, 1955–65, 1965–73, and 1973–78, indicates that the sectoral productivity growth rates have been dramatically changing. Some industries like communications have been growing steadily, while sectors like mining, public utilities, and construction have lost their pre-1965 expansion rates. In fact, in the period 1973–78, labor productivity growth in mining and construction has been negative, while transportation, wholesale trade, services, and public utilities productivity growth rates have slowed considerably.

Substantial changes in sectoral mix of employment and capital growth have occurred, particularly since the early 1970's. For example, the exodus of labor from farming has slowed, while mining and transportation have experienced substantial positive growth in employment. The employment growth rates have been slowed considerably in all sectors of the economy during 1974–78, except in mining and transportation. Deceleration of output growth is also evident in most industries for this period, with the exception of communications. This growth rate is even negative in construction and mining. Also, the growth of capital formation in various sectors has been reduced

*National Bureau of Economic Research and New York University. I am indebted to Irene Yew for her very able research assistance, and to Elliot Grossman for making his unpublished data series available to me. This research has been supported by the National Science Foundation, Grant No. 3301.

[1]For examples, see William Nordhaus, J. R. Norsworthy and Lawrence Fulco, and Lester Thurow.

substantially, and in some sectors, the slowdown is dramatic. Further, it is well documented that the growth of the stock of *R&D* has declined by about 2.5–3.0 percent in recent years.

These changes in sectoral growth rates of inputs and outputs over time have certainly affected productivity growth at the aggregate level. The results reported here indicate that the recent slowdown of productivity growth at both sectoral and aggregate levels can be explained to a considerable degree by the deceleration of output growth and inadequate utilization of existing capacity, plus the slowdown of growth of the capital-labor ratios, and the pace of technical change in some sectors.

II

The basic model of labor productivity growth is derived from an admittedly simple production function—a three-input Cobb-Douglas production function with neutral technical changes. The inputs are capital, labor, and stock of *R&D* for the period 1948–78. Following Nordhaus, I estimate the demand for each sector as a function of its price relative to the general price level, the difference between actual and normal unemployment rates, and the level of normal aggregate output, a measure calculated by George Perry. The estimated normal demand for a given sector is then calculated by setting the level of actual employment equal to its normal level. The production function relates normal output to levels of man-hours, stocks of fixed capital and *R&D* capital, and a linear time trend. By setting normal output and normal demand equal, assuming constant returns to scale in production, and that the adjustments of man-hours to changes in short-term demand depends on the level of capacity utilization and to its rate of change, the following productivity equation is obtained:

$$\begin{aligned} \ln P = {} & \alpha_0 + \alpha_1 \ln k + \alpha_2 \ln U_t \\ & + \alpha_2 \Delta \ln U_t + \alpha_4 \ln R + \alpha_5 t \end{aligned}$$

where P is the level of output per man-hour, k is the ratio of gross capital stock to man-hours, U is the gap between the rates of growth of actual and normal output, $\Delta \ln U_t$ is the change in $\ln U$, R is the stock of *R&D*, and t is the time trend. The variable R is a measure of the aggregate stock of *R&D* developed by John Kendrick, which was extended for the period 1969–78 using *NSF* published *R&D* data.

The productivity equation was fitted using the aggregate and sectoral data for 1949–78. The fit of the equations were very good, based on conventional criteria. However, prior to the discussion of the results, a few remarks about the problems of estimation seem to be in order. For some industries, we have restricted the capital-labor ratio and the utilization rate to the same coefficients. This was done once we observed that the two sets of coefficients were exactly the same when the variables were introduced separately. However, the most troublesome estimation problem arose in the sectoral equations due to the multicollinearity between the stock of *R&D* and the time trend. When the time trend was excluded from the regression, the coefficient of the *R&D* variable had the correct sign and significant magnitude in almost all of the sectoral equations. This is not surprising. Our measure of *R&D* refers to the aggregate level. What is needed to distinguish between the effects of disembodied technical change and the growth of *R&D* stock in specific sectoral productivity equations, however, is a measure of sector-specific stock of *R&D*. Unfortunately except for the period since 1958 for the manufacturing industries, the necessary data are not available. Another reason may be that if the appropriate measure of capital, that is, net capital stock, was used in estimating the model, the *R&D* variable could have been statistically significant. However, for the aggregate economy, the results indicate that we can distinguish between the effects of stock of *R&D* and disembodied technical change. The magnitudes of the coefficients of stock of *R&D* and the time trend were both statistically significant and positive. They were .06 and .01, respectively, in the aggregate economy productivity equations, and .10 and .01 in

the regression equation for the total private economy. These estimates are reasonable and consistent with some previous evidence.

My results also indicate that the level of productivity in both aggregate and industrial sectors are affected substantially by growth of the capital-labor ratios. These results are in contrast to those found by Nordhaus, Perry, and, most recently, Thurow, who did not find any effect of capital stock on growth of productivity. They are, however, consistent with results recently reported by Peter Clark. Evidently, the coefficients of the capital-labor ratio we have estimated are slightly greater than the share of capital in each sector or in the aggregate economy. This can be expected since our measure of capital is gross capital stock. The unavailability of net capital stock series for each sector for the entire 1948–78 period rendered the use of gross stock series necessary.

The coefficients of the short-term variables were also statistically significant in each equation, which implies that measured productivity growth is affected significantly by the level of utilization and its rate of change. In some industries, $\Delta \ln U_t$ was not significant, while in manufacturing industries, trade, construction, and public utilities, this variable was significant. A more systematic variable that appeared in every regression was the level of the utilization rate, suggesting that an important factor in restraining growth of measured productivity has been the failure of output to grow fast enough to catch up with its normal growth rate. Finally, the results indicated that disembodied technical change affects productivity growth positively and significantly in each sector except for the three industries which have been experiencing negative productivity growth since 1973—mining, construction, and public utilities.

III

Given the estimates, we can calculate the contribution of the capital-labor ratio, the short-term demand variables, and the *R&D* toward the slowdown of productivity growth since 1973 in the aggregate economy and its various sectors. These contributions toward the slowdown of productivity growth in period 1974–78 are shown in Table 1. Several aspects of these results are interesting. The slowdown of the capital-labor ratio growth, and the decline in the utilization rate and its rate of change contribute to the decline of the aggregate and sectoral productivity growth rates. The slowdown in growth of aggregate stock of *R&D* contributes about one-fourth and one-third of the slowdown of productivity in total economy, respectively. These estimates are probably on the high side and could be improved by further refinement of the data. However, they clearly indicate that slowdown of aggregate labor productivity can be explained fairly well by the conventional factors.

The contributions of these factors in explaining sectoral productivity growth are also significant. In most sectors, the entire or large percentage of productivity growth is due to slowdown in growth of the capital-labor ratios, degree of utilization, and the retardation of technical progress in some industries. In some sectors like farming and services, the slowdown of productivity growth rate is overexplained. In sectors where the growth or retardation of productivity growth has been substantial, for instance, mining, transportation, and public utilities, this set of factors account for a significant portion of the slowdown. In other sectors such as finance and communications, a substantial residual remains unexplained. It is possible that certain industry-specific factors like regulatory restriction on public utilities, or special factors in the construction industry, should be explicitly taken into account. Also, the strong growth of the communication sector needs special attention. I have not made such an effort at this point, but will do so in the future.

Note that the magnitudes of the contribution of growth of capital-labor ratios, the short-term utilization rate, and technical progress vary considerably among the various sectors. The estimates of the contribution of the capital-labor ratio are probably on the high side, as noted earlier, because I

TABLE 1—CONTRIBUTIONS TO THE SLOWDOWN OF AGGREGATE AND SECTORAL GROWTH RATES OF LABOR PRODUCTIVITY IN 1974–78

Industry	Changes in Growth Rate (1)	Capital-Labor Ratio (2)	Level of Utilization (3)	Changes in Level of Utilization (4)	Stock *R&D* (5)	Time Trend (6)	Total Cols. (2)–(6)
Total Economy	−0.93	−0.399	−0.439	−0.049	−0.18	0.01	−1.057
Total Private Economy	−0.78	−0.273	−0.222	−0.035	−0.30	0.01	−0.820
Farming	−1.17	−1.795	0.331			0.0004	−1.461
Mining	−6.79	−3.02	−0.576			−0.0008	−3.597
Construction	−1.06	−2.06	−0.004	0.440		−0.02	−1.644
Transportation	−2.45	−0.3996	−1.994			0.03	−2.364
Communication	2.52	0.248	0.211			0.074	0.533
Public Utilities	−2.66	0.041	−3.267	−0.117		−0.003	−3.346
Trade	−1.86	−1.215	−0.568	0.009		0.02	−1.750
Finance	1.59	−0.640	0.565			0.14	0.065
Services	−1.37	−1.37	−0.839			−0.004	−2.213
Government	−0.33	−1.64	−0.080			0.005	−0.239
Manufacturing	−1.02	−0.232	−0.662	−0.043		0.02	−0.917
Nondurables	−0.94	0.034	−1.015	−0.076		0.02	−1.037
Durables	−1.07	0.252	−0.762	−0.153		0.02	−0.643

have used gross capital stock measures, while the appropriate data are the net stock series. Also, note that the contribution of the capital-labor ratio and the other variables are not always in the same direction. This is particularly true for the growth rate of technical progress, which in all sectors except for mining, public utilities, and services, contributes positively to growth of productivity and offsets the negative effects exerted by the slow growth of the capital-labor ratio and the utilization rate.

The results shown in Table 1 are preliminary and can certainly be improved with better data and, perhaps better specification of the model. For example, we have not controlled adequately for the specific effects of the sharp recession of 1974, which some investigators found to have significantly affected productivity growth. Also, the estimates could be improved if adjustments were made for quality improvement in our capital and labor series. However, my results suggest that the slow growth of capital formation, the inability of the economy and various sectors to grow along their normal growth paths, and some slowdown in the rate of technical change, go a fair distance in explaining the slowdown of productivity growth since 1973.

High and steady growth of demand and rapid capital formation could lead to substantial improvement in aggregate and sectoral productivity growth. But the crucial policy decision is how to stimulate growth of demand and capital accumulation and at the same time lower the present high inflation.

REFERENCES

R. Brenner, "Technology, Labor and Economics," study no. 29, Data Resources, Inc., 1979.

P. K. Clark, "Capital Formation and the Recent Productivity Slowdown," mimeo., Council Economic Advisors, Dec. 20, 1977.

John W. Kendrick, *The Formation and Stocks of Total Capital*, New York 1976.

W. D. Nordhaus, "The Recent Productivity Slowdown," *Brookings Papers*, Washington 1972, *3*, 493–546.

J. R. Norsworthy and L. Fulco, "Productivity and Costs in the Private Economy," *Mon. Labor Rev.*, June 1976, *99*, 3–11.

G. L. Perry, "Potential Output and Productivity," *Brookings Papers*, Washington 1977, *1*, 11–47.

L. C. Thurow, "The U.S. Productivity Problem," mimeo., M.I.T., Apr. 1979.

DISCUSSION

JEROME A. MARK, U.S. Bureau of Labor Statistics: The problems Albert Rees mentions are certainly present in the productivity measures, and the government agencies, particularly the Bureau of Labor Statistics, have been aware of them. While I would agree that there probably is some downward bias to the *BLS* measures of productivity, I do not believe it is very large, as may seem to be indicated by Rees' observations.

The problem of inclusion of input measures for output in the National Income and Product Accounts has a very small impact on the *BLS* measure, which excludes general government, households, and institutions precisely for this reason. As Rees mentions, only 5 percent of the measure of total real output is probably affected by this problem.

On the price deflators, I would agree that the lag in the introduction of new and changing products like computers does lead to an upward bias in the price measures and hence a downward bias in the output measure. However, on the general question of quality change I am not sure that we know the overall direction of the quality error. Jack Triplett, in his survey of studies measuring quality change in the price indexes (in *Analysis of Inflation*, edited by Paul H. Earl), observed that some studies have produced quality-adjusted price indexes which rise more slowly than counterpart components in the *CPI* or *WPI* deflators, while for others the indexes have risen more than the comparable *BLS* or *BEA* indexes. Also, the Rees panel report pointed out that the price indexes can contain an upward or downward bias because of the measurement of quality change. As Triplett concludes, we are a long way from being able to estimate the quality error in the indexes or even to estimate the sign of the error in the overall indexes.

The use of hours paid rather than worked most likely results in a downward bias, but at present we only have a very crude estimate of its magnitude and that is not very large. The *BLS* task force did examine an hours-worked measure using various sources such as the surveys of employer expenditure for employee compensation, leave practice, and tenure data from area wage surveys. The best data from these surveys—some now discontinued—indicated that the adjustment of the hours-paid series to hours worked reduces the trend in aggregate hours for the nonfarm sector by 0.1 percent per year from 1952–75.

With respect to weighted labor input measures, *BLS* is now experimenting with the technique developed by Frank Gollop and Dale Jorgenson for weighting labor input by hourly compensation in age, sex, education, occupation, and industry categories. This method, which is similar to that developed by Ed Denison, rests on the assumption that categories of labor input are paid according to their respective marginal products. This assumption, as Rees points out, is tenuous in some cases.

The *BLS* welcomes the recommendation that multifactor productivity measurement be undertaken and research effort has already been directed along those lines. The undertaking is a major one, however. The issues of capital measurement are particularly troublesome, and capital measurement for productivity analysis rests on different principles from those appropriate to national income accounting. We are also working with the *BEA* and the Bureau of the Census to try to improve the data collection to permit better capital measurement. Key problems include more detailed measurement of investment by asset class, better information on retirements and depreciation of capital goods, information on land holdings, and detail on inventories by finished goods, raw materials, and goods in process categories.

At present, reasonably reliable data on energy inputs are available only for the manufacturing sector, and that only annually. Thus, multifactor productivity measures which include energy and materials can be constructed only for that sector. Similarly, capital stock data outside agriculture and manufacturing are not separated by industry on a basis consistently updated and

consistent with the National Income Accounts. Inclusion of intermediate goods and energy in multifactor productivity measure has been shown to be quite important. However, our initial efforts will cover only labor and capital outside the manufacturing sector.

The evident and recently worsening decline in productivity growth has added to the inflationary pressures in the economy, and many leaders in government, business, and labor have raised significant questions about the *BLS* productivity measurement program and its present scope. The Rees panel report has provided a sound basis on which we expect to improve and extend measurement and analysis of both aggregate and industry productivity.

EDWARD F. DENISON, Bureau of Economic Analysis: A reduction in *R&D*, or in the ratio of *R&D* expenditures to *GNP*, is frequently included in the litany of causes of the slowdown of growth since 1973. In considering this, I have usually started from Zvi Griliches' 1973 International Economic Association article, which contains what is probably the best calculation available for *R&D*'s contribution to growth before the slowdown. In it, Griliches estimated that as of about 1966, *R&D* contributed at most 0.3 percentage points to the growth rate of private domestic *GNP*, and probably considerably less.

A key ingredient of such calculations is the amount of *R&D* expenditures for projects that, if successful, can be expected to raise output per unit of input, given the way that output is actually measured. The characteristics of price indexes are such that, broadly speaking, only advances in knowledge that reduce the unit cost of providing types of final products that already exist raise measured output per unit of input. Only *R&D* directed either toward new processes, which may be roughly identified with research to reduce a firm's own costs, or toward new *intermediate* products and capital goods, can be expected to have this effect. But nearly all federally financed *R&D* and also most industry-financed *R&D* fail to meet this standard. Griliches counts only "effective" *R&D* in order to take this into account, but probably includes too much in his *R&D* series so his estimate of its contribution is likely to be overgenerous.

Starting from Griliches' estimate that *R&D* contributed at most 0.3 percentage points to the growth rate before the productivity slowdown, I concluded in *Accounting for Slower Economic Growth* that *R&D*'s contribution to the slowdown itself fell in the range from nothing at all to 0.1 points in the growth rate. Griliches reaches the same conclusion. He arrives at 0.1 percentage points "at most" as the decline for which *R&D* was responsible even though his procedures tend to overstate the amount because he uses the ratio of effective *R&D* to *GNP*, rather than its absolute amount; and because he assumes that the ratio of effective *R&D* to total *R&D* did not rise at a time when the share of defense and space *R&D* was shrinking.

I have been of the opinion that it is not feasible to estimate the contribution of *R&D* to growth by use of correlation analysis, which might be considered an alternative approach. If I understand Griliches correctly, he finds that recent years make it even more difficult than formerly to obtain any reliable results from correlation analysis.

My agreement with Griliches does not imply that everyone is of the same opinion. John Kendrick, by counting all *R&D* as pertinent to growth of measured output, arrives at a larger *R&D* contribution before the slowdown and blames *R&D* for somewhat more of the slowdown. Roger Brinner has the largest estimate of the contribution to the decline: 0.2 percentage points from 1965–70 to 1970–75, following a previous 0.2 points from 1960–65 to 1965–70. Brinner not only counts all *R&D*, but also deducts obsolescence from the stock of knowledge by a formula that relates it not to the new knowledge responsible for obsolescence but to the total stock of knowledge. But even Brinner's results, which I consider untenable, would explain little of the sharp drop in productivity that occurred after 1973.

The contribution to growth by *R&D* should not be identified with that of all advances in knowledge, which, my estimates suggest, contributed about 1.4 points to the 1948–73 growth rate in nonresidential business. Conceptually, this estimate covers organizational and managerial, as well as technological, knowledge. Technological knowledge results not only from organized *R&D*, but also from observation and experience of workers and managers, individual inventors, and all other sources. Finally, organized *R&D* throughout the world contributes to growth. Consequently, it is no great surprise to find that organized *R&D* conducted in the United States, as measured by *NSF*, is responsible for only a small fraction of the contribution of advances in knowledge—my estimates suggest less than one-fourth in 1948–73.

Transaction Cost Determinants of "Unfair" Contractual Arrangements

By BENJAMIN KLEIN*

Terms such as "unfair" are foreign to the economic model of voluntary exchange which implies anticipated gains to all transactors. However, much recent statutory, regulatory and antitrust activity has run counter to this economic paradigm of the efficiency properties of "freedom of contract." The growth of "dealer day in court" legislation, FTC franchise regulations, favorable judicial consideration of "unequal bargaining power," and unconscionability arguments, are some examples of the recent legal propensity to "protect" transactors. This is done by declaring unenforceable or illegal particular contractual provisions that, although voluntarily agreed upon in the face of significant competition, appear to be one-sided or unfair. Presentation of the standard abstract economic analysis of the mutual gains from voluntary exchange is unlikely to be an effective counterweight to this recent legal movement without an explicit attempt to provide a positive rationale for the presence of the particular unfair contractual term. This paper considers some transaction costs that might explain the voluntary adoption of contractual provisions such as termination at will and long-term exclusive dealing clauses that have been under legal attack.

I. The "Hold-up" Problem

In attempting to explain the complicated contractual details of actual market exchange, I start by noting that complete, fully contingent, costlessly enforceable contracts are not possible. This is a proposition obvious to even the most casual observer of economic phenomenon. Rather than the impersonal marketplace of costlessly enforceable contracts represented in standard economic analysis, individuals in most real world transactions are concerned with the possibility of breach and hence the identity and reputation of those with whom they deal. Further, even a cursory examination of actual contracts indicates that the relationship between transacting parties often cannot be fully described by a court-enforceable formal document that the parties have signed (see Stewart Macauley). While the common law of contracts supplies a body of rules and principles which are read into each contract, in many cases explicit terms (which include these general unwritten terms) remain somewhat vague and incomplete.

Contracts are incomplete for two main reasons. First, uncertainty implies the existence of a large number of possible contingencies and it may be very costly to know and specify in advance responses to all of these possibilities. Second, particular contractual performance, such as the level of energy an employee devotes to a complex task, may be very costly to measure. Therefore contractual breach may often be difficult to prove to the satisfaction of a third-party enforcer such as a court.

Given the presence of incomplete contractual arrangements, wealth-maximizing transactors have the ability and often the incentive to renege on the transaction by holding up the other party, in the sense of taking advantage of unspecified or unenforceable elements of the contractual relationship. Such behavior is, by definition,

*Professor of economics, University of California-Los Angeles. This paper was written while I was a Law and Economics Fellow at the University of Chicago Law School. Armen Alchian, Roy Kenney, Edmund Kitch, Timothy Muris, Richard Posner, and George Priest provided useful comments on earlier drafts.

unanticipated and not a long-run equilibrium phenomenon. Oliver Williamson has identified and discussed this phenomenon of "opportunistic behavior," and my recent paper with Robert Crawford and Armen Alchian attempted to make operational some of the conditions under which this hold-up potential is likely to be large. In addition to contract costs, and therefore the incompleteness of the explicit contract, we emphasized the presence of appropriable quasi rents due to highly firm-specific investments. After a firm invests in an asset with a low-salvage value and a quasi-rent stream highly dependent upon some other asset, the owner of the other asset has the potential to hold up by appropriating the quasi-rent stream. For example, one would not build a house on land rented for a short term. After the rental agreement expires, the landowner could raise the rental price to reflect the costs of moving the house to another lot.[1]

The solution we emphasized was vertical integration, that is, one party owning both assets (the house and the land). Because the market for land is competitive, the price paid for the land by the homebuilder does not reflect these potentially appropriable quasi rents. However, this solution will not necessarily be observed. The size of the hold-up potential is a multiplicative function of two factors: the presence of specific capital, that is, appropriable quasi rents, and the cost of contractually specifying and enforcing delivery of the service in question —the incentive for contract violation and the ease of contract violation. Even where there is a large amount of highly specific capital, the performance in question may be cheaply specifiable and measureable and a complete contract legally enforceable at low cost. Therefore, while a short-term rental contract is not feasible, a possible solution may be a long-term lease. In addition, since the cases we will be considering deal with human capital, vertical integration in the sense of outright ownership is not possible.

II. Contractual Solutions

Since the magnitude of the potential holdup may be anticipated, the party to be cheated can merely decrease the initial price he will pay by the amount of the appropriable quasi rents. For example, if an employer knows that an employee will cheat a certain amount each period, it will be reflected in the employee's wage. Contracts can be usefully thought to refer to anticipated rather than stated performance. Therefore the employee's behavior should not even be considered "cheating." A secretary, for example, may miss work one day a week on average. If secretary time is highly substitutable, the employer can cut the secretary's weekly wage 20 percent, hire 20 percent more secretaries and be indifferent. The secretary, on the other hand, presumably values the leisure more than the additional income and therefore is better off. Rather than cheating, we have a voluntarily determined, utility-maximizing contractual relationship.

In many cases, however, letting the party cheat and discounting his wage will not be an economical solution because the gain to the cheater and therefore his acceptable compensating wage discount is less than the cost to the firm from the cheating behavior. For example, it is easy to imagine many cases where a shirking manager will impose costs on the firm much greater than his personal gains. Therefore the stockholders cannot be made indifferent to this behavior by cutting his salary and hiring more lazy managers. The general point is that there may not be perfect substitutability between quantity and quality of particular services. Hence, even if one knew that an unspecified

[1]This problem is different from the standard monopoly or bilateral monopoly problem for two reasons. First, market power is created only after the house investment is made on a particular piece of land. Such postinvestment power can therefore exist in many situations that are purely competitive preinvestment. Second, the problem we are discussing deals with the difficulties of contract enforcement. Even if some preinvestment monopoly power exists (for example, a union supplier of labor services to harvest a crop), if one can write an enforceable contract preinvestment (i.e., before the planting), the present discounted value of the monopoly return may be significantly less than the one-time postinvestment hold-up potential (which may equal the entire value of a crop ready to be harvested).

element of quality would be reduced by a certain amount in attempting the holdup, an *ex ante* compensatory discount in the quoted price of the promised high quality service to the cost of providing the anticipated lower-quality supply would not make the demander of the service indifferent. Individuals would be willing to expend real resources to set up contractual arrangements to prevent such opportunism and assure high-quality supply.

The question then becomes how much of the hold-up problem can be avoided by an explicit government-enforced contract, and how much remains to be handled by an implicit self-enforcing contract. This latter type of contract is one where opportunistic behavior is prevented by the threat of termination of the business relationship rather than by the threat of litigation. A transactor will not cheat if the expected present discounted value of quasi rents he is earning from a relationship is greater than the immediate hold-up wealth gain. The capital loss that can be imposed on the potential cheater by the withdrawal of expected future business is then sufficient to deter cheating.

In our forthcoming article, Keith Leffler and I develop this market-enforcement mechanism in detail. It is demonstrated that one way in which the future-promised rewards necessary to prevent cheating can be arranged is by the payment of a sufficiently high-price "premium." This premium stream can usefully be thought of as "protection money" paid to assure noncheating behavior. The magnitude of this price premium will be related to the potential holdup, that is, to the extent of contractual incompleteness and the degree of specific capital present. In equilibrium, the present discounted value of the price-premium stream will be exactly equal to the appropriable quasi rents, making the potential cheater indifferent between cheating and not. But the individual paying the premium will be in a preferable position as long as the differential consumer's surplus from high-quality (noncheating) supply is greater than the premium.

One method by which this equilibrium quasi-rent stream can be achieved without the existence of positive firm profits is by having the potential cheater put up a forfeitable-at-will collateral bond equal to the discounted value of the premium stream. Alternatively, the potential cheater may make a highly firm-specific productive investment which will have only a low-salvage value if he cheats and loses future business. The gap between price and salvageable capital costs is analytically equivalent to a premium stream with the nonsalvageable asset analytically equivalent to a forfeitable collateral bond.

III. "Unfair" Contractual Terms

Most actual contractual arrangements consist of a combination of explicit- and implicit-enforcement mechanisms. Some elements of performance will be specified and enforced by third-party sanctions. The residual elements of performance will be enforced without invoking the power of some outside party to the transaction but merely by the threat of termination of the transactional relationship. The details of any particular contract will consist of forms of these general elements chosen to minimize transaction costs (for example, hiring lawyers to discover contingencies and draft explicit terms, paying quality-assurance premiums, and investing in nonsalvageable "brand name" assets) and may imply the existence of what appears to be unfair contract terms.

Consider, for example, the initial capital requirements and termination provisions common in most franchise contractual arrangements. These apparently one-sided terms may be crucial elements of minimum-cost quality-policing arrangements. Given the difficulty of explicitly specifying and enforcing contractually every element of quality to be supplied by a franchisee, there is an incentive for an individual opportunistic franchisee to cheat the franchisor by supplying a lower quality of product than contracted for. Because the franchisee uses a common trademark, this behavior depre-

ciates the reputation and hence the future profit stream of the franchisor.[2]

The franchisor knows, given his direct policing and monitoring expenditures, the expected profit that a franchisee can obtain by cheating. For example, given the number of inspectors hired, he knows the expected time to detect a cheater; given the costs of low-quality inputs he knows the expected extra short-run cheating profit that can be earned. Therefore the franchisor may require an initial lump sum payment from the franchisee equal to this estimated short-run gain from cheating. This is equivalent to a collateral bond forfeitable at the will of the franchisor. The franchisee will earn a normal rate of return on that bond if he does not cheat, but it will be forfeited if he does cheat and is terminated.

In many cases franchisee noncheating rewards may be increased and short-run cheating profits decreased (and therefore franchisor direct policing costs reduced) by the grant of an exclusive territory or the enforcement of minimum resale price restraints (see my paper with Andrew McLaughlin). Franchisors can also assure quality by requiring franchisee investments in specific (nonfully salvageable) production assets that upon termination imply a capital-cost penalty larger than any short-run wealth gain that can be obtained by the franchisee if he cheats. For example, the franchisor may require franchisees to rent from them short term (rather than own) the land upon which their outlet is located. This lease arrangement creates a situation where termination implies that the franchisor can require the franchisee to move and thereby impose a capital loss on him up to the amount of his initial nonsalvageable investment. Hence a form of collateral to deter franchisee cheating is created.[3]

It is important to recognize that franchise termination, if it is to assure quality compliance on the part of franchisees, must be unfair in the sense that the capital cost imposed on the franchisee that will optimally prevent cheating must be larger than the gain to the franchisee from cheating. Given that less than infinite resources are spent by the franchisor to monitor quality, there is some probability that franchisee cheating will go undetected. Therefore termination must become equivalent to a criminal-type sanction. Rather than the usually analyzed case of costlessly detected and policed contract breach, where the remedy of making the breaching party pay the cost of the damages of his specific breach makes economic sense, the sanction here must be large enough to make the expected net gain from cheating equal to zero. The transacting parties contractually agree upon a penalty-type sanction for breach as a means of economizing on direct policing costs. Because contract enforcement costs (including litigation costs which generally are not collectable by the innocent party in the United States) are not zero, this analysis provides a rationale against the common law prohibition of penalty clauses.

The obvious concern with such seemingly unfair contractual arrangements is the possibility that the franchisor may engage in opportunistic behavior by terminating a franchisee without cause, claiming the franchise fee and purchasing the initial franchisee investment at a distress price. Such behavior may be prevented by the

[2]At locations where this incentive is very large, for example, on superhighways where the probability of repeat sales by particular customers is very low, the franchisor may "vertically integrate" and not compensate their employees on any profit-sharing basis. Such fixed wage compensation schemes reduce the incentive to cheat but at the cost of reducing the incentive for workers to supply any effort that is not explicitly specified and measureable by the employer. It is this latter incentive that is harnessed by franchising arrangements.

[3]The initial franchise investment also serves as a means of establishing an efficient compensation mechanism. Because the franchise investment is a saleable asset it provides a market measure of future profit and hence a precise incentive on franchisee efforts to build up the business. While an employee contract can contain a profit-sharing arrangement, and retirement and stock option provisions to reward employee efforts that yield a return far in the future and protect the employee's heirs, it would be extremely difficult to write *ex ante* complete, enforceable (i.e., measureable) contract terms that would as accurately reflect the value of marginal employee efforts.

depreciation of the franchisor's brand name and therefore decreased future demand by potential franchisees to join the arrangement. However, this protective mechanism is limited by the relative importance of new franchise sales compared to the continuing franchising operation, that is, by the "maturity" of the franchise chain.

More importantly, what limits reverse cheating by franchisors is the possible increased cost of operating the chain through an employee operation compared to a franchise operation when such cheating is communicated among franchisees. As long as the implicit collateral bond put up by the franchisee is less than the present discounted value of this cost difference, franchisor cheating will be deterred. Although explicit bonds and price premium payments cannot simultaneously be made by both the franchisee and the franchisor, the discounted value of the cost difference has the effect of a collateral bond put up by the franchisor to assure his noncheating behavior. This explains why the franchisor does not increase the initial franchise fee to an arbitrarily high level and correspondingly decrease its direct policing expenditures and the probability of detecting franchisee cheating. While such offsetting changes could continue to optimally deter franchisee cheating and save the real resource cost of direct policing, the profit from and hence the incentive for reverse franchisor cheating would become too great for the arrangement to be stable.

Franchisees voluntarily signing these agreements obviously understand the termination-at-will clause separate from the legal consequences of that term to mean nonopportunistic franchisor termination. But this does not imply that the court should judge each termination on these unwritten but understood contract terms and attempt to determine if franchisor cheating has occurred. Franchisees also must recognize that by signing these agreements they are relying on the implicit market-enforcement mechanism outlined above, and not the court to prevent franchisor cheating. It is costly to use the court to regulate these terminations because elements of performance are difficult to contractually specify and to measure. In addition, litigation is costly and time consuming, during which the brand name of the franchisor can be depreciated further. If these costs were not large and the court could cheaply and quickly determine when franchisor cheating had occurred, the competitive process regarding the establishment of contract terms would lead transactors to settle on explicit governmentally enforceable contracts rather than rely on this implicit market-enforcement mechanism.

The potential error here is, after recognizing the importance of transaction costs and the incomplete "relational" nature of most real world contracts, to rely too strongly on the government as a regulator of unspecified terms (see Victor Goldberg). While it is important for economic theory to handle significant contract costs and incomplete explicit contractual arrangements, such complexity does not imply a broad role for government. Rather, all that is implied is a role for brand names and the corresponding implicit market enforcement mechanism I have outlined.

IV. Unequal Bargaining Power

An argument made against contract provisions such as termination-at-will clauses is that they appear to favor one party at the expense of another. Hence it is alleged that the terms of the agreement must have been reached under conditions of "unequal bargaining power" and therefore should be invalid. However, a further implication of the above analysis is that when both parties can cheat, explicit contractual restraints are often placed on the smaller, less well-established party (the franchisee), while an implicit brand name contract-enforcement mechanism is relied on to prevent cheating by the larger, more well-established party (the franchisor).

If information regarding quality of a product supplied by a large firm is communicated among many small buyers who do not all purchase simultaneously, the potential holdup relative to, say, annual sales is reduced substantially compared to

the case where each buyer purchased from a separate independent small firm. There are likely to be economies of scale in the supply of a business brand name, because in effect the large firm's total brand name capital is put on the line with each individual sale. This implies a lower cost of using the implicit contract mechanism, that is, a lower-price premium necessary to assure non-breach, for a large firm compared to a small firm. Therefore one side of the contract will be relatively more incomplete.

For example, in a recent English case using the doctrine of inequality of bargaining power to bar contract enforcement, an individual songwriter signed a long-term (ten-year) exclusive service contract with a music publisher for an agreed royalty percentage.[4] Since it would be extremely costly to write a complete explicit contract for the supply of publishing services (including advertising and other promotion activities whose effects are felt over time and are difficult to measure), after a songwriter becomes established he has an incentive to take advantage of any initial investment made by a publishing firm and shift to another publisher. Rather than rely on the brand name of the songwriter or require him to make a specific investment which can serve as collateral, the exclusive services contract prevents this cheating from occurring.

The major cost of such explicit long-term contractual arrangements is the rigidity that is created by the necessity of setting a price or a price formula *ex ante*. In this song publishing case, the royalty formula may turn out *ex post* to imply too low a price to the songwriter (if, say, his cooperative promotional input is greater than originally anticipated.) If the publisher is concerned about his reputation, these royalty terms will be renegotiated, a common occurrence in continuing business relationships.

If an individual songwriter is a small part of a large publisher's total sales, and if the value of an individual songwriter's ability generally depreciates rapidly or does not persist at peak levels so that signing up new songwriters is an important element of a publisher's continuing business, then cheating an individual songwriter or even all songwriters currently under contract by refusing to renegotiate royalty rates will imply a large capital cost to the publisher. When this behavior is communicated to other actual or potential composers, the publisher's reputation will depreciate and future business will be lost. An individual songwriter, on the other hand, does not generally have large, diversified long-term business concerns and therefore cannot be penalized in that way. It is therefore obvious, independent of any appeal to disparity of bargaining power, why the smaller party would be willing to be bound by an explicit long-term contract while the larger party is bound only implicitly and renegotiates terms that turn out *ex post* to be truly divergent from *ex ante*, but unspecified, anticipations.

However, the possibility of reverse publisher cheating is real. If, for example, the songwriter unexpectedly becomes such a great success that current sales by this one customer represents a large share of the present discounted value of total publisher sales, the implicit contract enforcement mechanism may not work. Individuals knowingly trade off these costs of explicit and implicit-enforcement mechanisms in settling upon transaction cost-minimizing contract terms. Although it would be too costly in a stochastic world to attempt to set up an arrangement where no cheating occurs, it is naive to think that courts can cheaply intervene to discover and "fix up" the few cases of opportunistic behavior that will occur. In any event, my analysis makes it clear that one cannot merely look at the agreed upon, seemingly "unfair" terms to determine if opportunism is occurring.

V. Conclusion

Ronald Coase's fundamental insight defined the problem. With zero transaction costs, the equilibrium form of economic organization is indeterminate. However, rather than distinguishing between the crude alternatives of vertical integration and

[4]See *Macaulay v. Schroeder Publishing Co., Ltd.* discussed in M. J. Trebilcock.

market exchange, what we really have to explain are different types of market-determined contractual relationships. I have argued that a particular form of transaction cost based upon the existence of incomplete contracts (due to uncertainty and measurement costs)—a transaction cost I have called the hold-up problem—may be an important reason in many cases for termination-at-will and exclusive-dealing contractual arrangements.

The danger is that a discussion of hold-up-type transaction costs can lead to *ad hoc* theorizing. The discussion here was meant to be suggestive. If economists are to explain satisfactorily the form of particular complex contracts adopted in the marketplace, they must "get their hands dirty" by closely investigating the facts and state of the law to determine hold-up possibilities and contract enforcement difficulties in particular cases. The most useful legal input to obtain knowledge of the institutional constraints on the trading process, is not likely to come from professors of contract law. Rather, we should consider the knowledge accumulated by practicing attorneys familiar with the likely hold-up problems and the contractual solutions commonly adopted in particular industries. When all firms in a particular industry use similar contractual provisions, it is unlikely to be the result of duress or fraud and should not necessarily be considered (as some courts have) as evidence of collusion. Such uniformity suggests the existence of independent attempts within a competitive environment to solve an important common problem and signals the presence of a prime research prospect.

REFERENCES

R. J. Coase, "The Nature of the Firm," *Economica*, Nov. 1937, *4*, 386–405.

V. P. Goldberg, "Toward an Expanded Economic Theory of Contract," *J. Econ. Issues*, Mar. 1976, *10*, 45–61.

B. Klein, R. G. Crawford, and A. A. Alchian, "Vertical Integration, Appropriable Rents and the Competitive Contracting Process," *J. Law Econ.*, Oct. 1978, *21*, 297–326.

______ **and K. Leffler,** "Non-Governmental Enforcement of Contracts: The Role of Market Forces in Guaranteeing Quality," *J. Polit. Econ.*, forthcoming.

______ **and A. McLaughlin,** "Resale Price Maintenance, Exclusive Territories and Franchise Termination: The Coors Case," unpublished manuscript, Univ. California-Los Angeles 1979.

S. Macauley, "Non-Contractual Relations in Business: A Preliminary Study," *Amer. Soc. Rev.*, Feb. 1963, *28*, 55–69.

M. J. Trebilcock, "The Doctrine of Inequality of Bargaining Power: Post-Benthamite Economics in the House of Lords," *Univ. Toronto Law J.*, Fall 1976, *26*, 359–85.

Oliver E. Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York 1975.

Strict Liability vs. Negligence in a Market Setting

By A. MITCHELL POLINSKY*

There are two competing rules of liability for controlling activities like speeding or polluting which cause harm to others. The rule of strict liability shifts the victim's harm to the injurer regardless of the injurer's behavior, while the rule of negligence shifts the burden to the injurer only if the injurer does not take some specified amount of care.

Most formal analyses of strict liability and negligence have been in a nonmarket context like automobile accidents.[1] These studies conclude (or imply) that both rules are efficient when the administering authority has sufficient information to set the negligence standard properly (in each case a defense of contributory negligence may be required). According to these analyses, strict liability is efficient because it fully internalizes the harm, and negligence is efficient because the injurer can be induced to take exactly the specified amount of care and this amount can be set efficiently. For reasons which will become apparent, it is important to note that these models have assumed a fixed number of injurers—for example, a predetermined number of drivers.

Interestingly, most informal discussions of strict liability and negligence have suggested that in a market setting negligence may be inefficient even when the administering authority has perfect information. The argument is stated clearly in a discussion of environmental control by Richard Stewart and James Krier:

> If strict liability is not imposed for the residual damages caused by partially controlled polluting activity, these damages will not be reflected in the price of commodities produced by such activity. As a result, commodities with whose production pollution is associated will be underpriced relative to commodities whose production causes no pollution, resulting in resource misallocation. [p. 227][2]

This paper formally analyzes strict liability and negligence in a market setting. The discussion emphasizes the impact of the rules on the market price and on the number of firms in the industry. For simplicity, the damage caused by each firm is assumed to be determined only by that firm's "care" (and not also by the firm's output or the victim's behavior).[3]

The argument that both strict liability and negligence are efficient is correct in the short run when the number of firms causing harm is fixed. The market price in the short run is the same under strict liability and negligence despite the fact that under the rule of negligence firms do not bear the cost of their harmful activity. The argument that only strict liability is efficient is correct in the long run when the number of firms is variable. When the negligence standard corresponds to the efficient level of care, the market price in the long run under negligence is too low and too many firms enter the industry. However, when the standard is

*Law School and economics department, Stanford University; and National Bureau of Economic Research. Work on this paper was supported by the National Science Foundation through a grant (SOC 78-20159) to the law and economics program of the National Bureau of Economic Research. Any opinions expressed are my own and not those of the NBER. Helpful comments were received from participants in a number of seminars.

[1]See, for example, the paper by John Brown. The only exceptions are discussed in fn. 4 below. There are also a number of formal studies of products liability in a market context which do not compare strict liability to negligence.

[2]See also the discussions by Guido Calabresi (pp. 500–17), Duncan Kennedy (pp. 54–61), and Richard Posner (pp. 137–42). Michael Spence and Martin Weitzman (pp. 216–17) make a similar point in the context of comparing taxes to standards.

[3]See comments (e) and (f) in Section IV below for a discussion of how the results are affected if harm also depends on output or the victim's behavior.

chosen optimally, taking into account the inefficiency of the negligence rule, the standard exceeds the efficient level of care, the price rises, and the number of firms falls. The optimal second best standard may actually result in the same price and number of firms under negligence as under strict liability.[4]

I. The Model

The analysis of strict liability and negligence is undertaken in a very simple partial-equilibrium model of a competitive industry composed of n identical firms. Each firm's cost of producing q units of output is $C(q)$. The average cost function is assumed to be U-shaped. Each firm's cost of taking z units of care is z (care is defined so that one unit costs one dollar). Associated with each firm's level of care is an external damage $D(z)$. Damages decrease with care. Let $P(s)$ be the industry's inverse demand function, where $s=nq$ is aggregate output. Social welfare W equals the benefits to consumers of output nq (assumed to be approximated by the area under the industry's inverse demand curve), less the cost of producing nq, the external damages, and the cost of taking care:

$$(1)\quad W=\int_0^{nq} P(s)\,ds-nC(q)-nD(z)-nz$$

In the short run, the number of firms is fixed.[5] The social optimum is defined by the q and z which maximize social welfare. From (1) the first-order condition (all solutions are assumed to be unique interior local optima) with respect to q may be written as

$$(2)\quad P(nq)=C'(q)$$

[4]A recent paper by Čento Veljanovski (pp. 17–22) uses a diagrammatic model to show that negligence is inefficient in the long run, but he does not consider the second best problem. The paper by Steven Shavell develops some closely related issues (see comment (e) in Section IV below). Barry Weingast, Mark McBride, and John Conant compare strict liability and negligence when, in effect, only the consumer of a product affects the harm.

[5]If the number of firms is effectively variable in the short run due to exit, the results may be different (see comment (d) in Section IV below).

This has the usual interpretation that price equals marginal cost. The first-order condition with respect to z is

$$(3)\quad -D'(z)=1$$

This states that the marginal benefits from an increase in care—reduced damages—equals the marginal cost of greater care.

In the long run, the number of firms is variable. The social optimum is now defined by the q, z, and n which maximize social welfare. From (1) the first-order condition with respect to n can be written as

$$(4)\quad P(nq)=\left[C(q)+D(z)+z\right]/q$$

This has the usual interpretation that price equals average cost.

The social welfare-maximizing values of the variables will be indicated by an asterisk.

II. Short-Run Analysis

Under strict liability, each firm's problem is to maximize revenues less production costs, external damages, and the cost of taking care:

$$(5)\quad \max_{q,z}\ Pq-C(q)-D(z)-z$$

The first-order conditions with respect to q and z are, respectively,

$$(6)\quad P=C'(q)$$

$$(7)\quad -D'(z)=1$$

which coincide with the social optimum. This is not surprising since strict liability fully internalizes the externality.

Under negligence, let $\bar{z}$ be the standard of care. Each firm's problem is

$$(8)\quad \max_{q,z}\ Pq-C(q)-\bar{D}(z)-z$$

where $$\bar{D}(z)=\begin{cases}0 & \text{if } z\geqslant\bar{z},\\ D(z) & \text{if } z<\bar{z}\end{cases}$$

The first-order condition with respect to q is the same as (6). It is easily shown that when

$\bar{z}=z^*$, the level of care chosen by each firm is $\bar{z}$ (see fn. 8 below). Thus, the social optimum can be reached under negligence if the standard is chosen properly, equal to the efficient level of care.

III. Long-Run Analysis

Under strict liability, each firm chooses output and care to maximize profits as indicated by (5), so that (6) and (7) are satisfied (implying at least that $z=z^*$). In addition, equilibrium is characterized by zero profits:

$$(9) \qquad P(nq)q-C(q)-D(z^*)-z^*=0$$

This is equivalent to the condition for the social optimum that price equals average cost. Thus, strict liability is efficient in the long run.

Under negligence, assume initially that the standard is set at the efficient level of care. Firms maximize profits by choosing output and care according to (8), so that (6) is satisfied and $z=z^*$. The zero profit condition under negligence is, therefore,

$$(10) \qquad P(nq)q-C(q)-z^*=0$$

This condition, along with (6), implies that the output of each firm under negligence is too low, that the market price is too low, and that too many firms enter the industry.[6]

Since the negligence outcome is not efficient when the standard equals the efficient level of care, the optimal second best standard is of some interest. It can be demonstrated that, optimally chosen, $\bar{z}>z^*$.[7] The reason for this is not hard to understand. Consider a slight increase in the standard starting at $\bar{z}=z^*$. Since firms will continue to meet the standard (they will over some range), their costs will rise and, as a consequence, so will the market price. This has the beneficial effect of raising the output of each firm (which was too low when $\bar{z}=z^*$) and of reducing the number of firms. The inefficiency created by inducing firms to take care just above the efficient level is negligible since the effect on social welfare of a marginal change in firms' care is zero at the optimal level of care. Thus, it is desirable to increase the standard to some extent.

There is a limit to how high the standard can be set and still induce firms to meet the standard. It is easy to show that the highest feasible standard leads to the same price under negligence as under strict liability; given this standard and the resulting price, the output of each firm and the number of firms is efficient under negligence.[8] By reasoning similar to that used in the previous paragraph, it might seem, therefore, that the optimal second best standard should not be this high: the effect on social welfare from lowering the level of care slightly (by lowering the standard) would be beneficial, while the effect on social welfare due to the resulting marginal changes in firm output and the number of firms would seem to be negligible. Perhaps surprisingly, this argument is not correct. It can be demonstrated that the optimal second best standard under negligence might in fact lead to the same price, firm output, and number of firms as under strict liability.[9]

[6]Let q^0 and n^0 be the equilibrium values of q and n under negligence. Substituting $P(nq)=C'(q)$ into (9) and (10) gives expressions of the form $C'(q)q-C(q)-K=0$, where K is some constant. It follows from $C''>0$ (rising marginal costs are assumed) that $dq/dK>0$. Thus, $q^0<q^*$. Since $P(nq)=C'(q)$ and $C''>0$, $P(n^0q^0)<P(n^*q^*)$. Thus, $n^0q^0>n^*q^*$ and $n^0>n^*$.

[7]Firms will just meet the standard if and only if $\bar{z}$ is between zero and some $\hat{z}>z^*$ (see fn. 8 below). Given $\bar{z}$ in this range, it follows from substituting (6) into (10) (with $\bar{z}$ replacing z^*) that the equilibrium q minimizes $(C(q)+\bar{z})/q$. Let $q(\bar{z})$ be the solution to this problem. Given $q(\bar{z})$, the price is $(C(q(\bar{z}))+\bar{z})/q(\bar{z})$. Thus, $nq(\bar{z})=P^{-1}[(C(q(\bar{z}))+\bar{z})/q(\bar{z})]$. Let $n(\bar{z})$ be the solution of this expression for n. Substituting $q(\bar{z})$ and $n(\bar{z})$ into (1) gives social welfare solely as a function of the standard, $W(\bar{q})$. After some manipulation, it can be seen that $\text{sgn}\{W'\}=\text{sgn}\{-n'D-n(D'+1)\}$. Since $n'<0$ it follows from assuming that $D(z)+z$ is U-shaped that $W'(\bar{z})>0$ for $\bar{z}<z^*$.

[8]If $\bar{z}<\hat{z}$, where $\hat{z}=D(z^*)+z^*$, firms will choose $\bar{z}$ because $\bar{z}<z+D(z)$ for all $z<\bar{z}$. If $\bar{z}=\hat{z}$, firms are indifferent between $\bar{z}$ and z^*. If $\bar{z}>\hat{z}$, firms will choose z^*. Thus, the highest feasible standard is $\hat{z}$ (or arbitrarily close to it). Given $\bar{z}=\hat{z}$, it follows from the discussion in fn. 7 above that $q(\hat{z})=q^*$ and that the price is $(C(q^*)+D(z^*)+z^*)/q^*$; since q and $P(nq)$ are the same as under strict liability, n must also be the same.

[9]It is sufficient to show that it is possible that $W'(\bar{z})>0$ for all $\bar{z}<\hat{z}$. Recall from fn. 7 above that $\text{sgn}\{W'\}=\text{sgn}\{-n'D-n(D'+1)\}$. Since $D'(z)$, $n(z)$, and $n'(z)$ are continuous functions on $0<z<\hat{z}$, they are bounded. Thus, if $D(\hat{z})$ is sufficiently large, the result follows.

The basis for this conclusion can be explained as follows. It is true that lowering the standard slightly from its highest feasible level improves social welfare through its effect on the level of care and leaves social welfare unchanged by its effect on firm output. It is not true, however, that lowering the standard leaves social welfare unchanged through its effect on the number of firms even though the number of firms was previously efficient. This is because the optimal number of firms *given* an inefficiently high level of care is less than the efficient number of firms. Since lowering the standard reduces firms' costs and consequently increases the number of firms, this effect on social welfare is undesirable. Therefore, whether the standard should be lowered depends on the strength of this effect relative to the beneficial effect on care.

IV. Discussion

(a) The model used here can be interpreted in two ways. The most natural interpretation is that the damage from each firm's output harms individuals who are not consumers of that output (for example, a chemical industry pollutes a lake used for recreation). The other interpretation is that the damage falls on the consumers of the firm's output but they underestimate it (for example, food producers sell products with misperceived carcinogenic effects). Under the second interpretation the model applies as it stands if the consumers are totally ignorant of the harm. However, it is easy to modify it to allow the demand curve for the output to shift, depending on who bears the damage. As long as consumers underestimate the damages, the basic results will apply.

(b) In a sense, the problem discussed here with the negligence rule in a market setting is that it is not possible to apply the rule to the industry as a whole. Courts are allowed to determine only whether any particular firm is negligent with respect to the activities over which it has control. This is not sufficient to correctly regulate the number of firms in the long run since all firms are acting the same way. No firm can be held responsible for the number of firms in the industry.

(c) The same type of observation would apply in a nonmarket setting in which there is free entry into the injuring activity. For example, consider individuals who must decide whether to drive to a nearby city—and thereby risk hitting innocent pedestrians —or to not make the trip at all. Under a negligence rule, too many individuals will make the trip even if all who do so take the proper amount of care in driving. The earlier formal analyses of strict liability and negligence in a nonmarket context which concluded that both rules were efficient assumed a fixed number of injurers.

(d) A version of the entry problem with the negligence rule also can arise in a market setting in the short run. For example, suppose there are a certain number of firms operating in the short run in the absence of any regulation and that these firms have different costs of production (but, for simplicity, the same damage functions and cost of taking care). The rule of strict liability—which would still be efficient with this generalization—may drive some of the costlier firms out of business. The rule of negligence, however, may induce some firms which would have gone out of business under strict liability to stay in business. As in the long run, this will result in too low a price and too many firms.[10]

(e) The main point of this paper would not be affected if each firm's total damage depends both on that firm's level of care and its output. If the negligence standard is then appropriately defined both with respect to the firm's care level and its output, the present analysis would apply essentially unchanged. If, however, the negligence standard is still defined just in terms of care (as is usual in practice), there would be an additional problem with the negligence rule besides the entry problem: it would distort the firm's choice between care and output.[11]

[10]This point is illustrated diagrammatically in the context of comparing taxes to subsidies in my 1979 paper.

[11]See Steven Shavell's paper. Since he treats the industry as composed of one competitively behaving

(f) Some of the conclusions of this paper may be affected if the total damage depends both on the firm's actions and the victim's behavior (for example, the care exercised in using a lawn mower). If the number of potential victims is fixed—so there is no "entry issue" on the victim's side—then the present results would be essentially unchanged (putting aside the separate issue discussed in the previous comment). An appropriate defense of contributory negligence—in which the firm would be free of liability if the victim did not take his efficient amount of care—should in principle be added to both the strict liability and negligence rules. In equilibrium the victim would take the correct amount of care, so the strict liability and negligence rules would operate as described in this paper. If, however, the number of victims is variable, then in general neither liability rule (with a defense of contributory negligence) would be expected to be efficient. Negligence would lead to excessive entry on the injurer side, while strict liability would encourage too much entry on the victim side.[12]

firm whose harm depends on care and output, he does not raise the problem of controlling the number of firms.

[12]An analogous result is shown by Shavell when the damage depends on the levels of care and activity of one injurer and one victim. The same point should apply when damages depend on care and the number of injurers and victims.

REFERENCES

J. P. Brown, "Toward an Economic Theory of Liability," *J. Legal Stud.*, June 1973, *2*, 323–49.

G. Calabresi, "Some Thoughts on Risk Distribution and the Law of Torts," *Yale Law J.*, Mar. 1961, *70*, 499–553.

D. Kennedy, "A History of Law and Economics, or the Fetishism of Commodities," unpublished paper, Harvard Univ., Sept. 1978.

A. M. Polinsky, "Notes on the Symmetry of Taxes and Subsidies in Pollution Control," *Can. J. Econ.*, Feb. 1979, *12*, 75–83.

Richard A. Posner, *Economic Analysis of Law*, 2d ed., Boston 1977.

S. Shavell, "Strict Liability versus Negligence," *J. Legal Stud.*, Jan. 1980, *9*, 1–25.

A. M. Spence and M. L. Weitzman, "Regulatory Strategies for Pollution Control," in Ann F. Friedlaender, ed., *Approaches to Controlling Air Pollution*, Cambridge 1978.

Richard B. Stewart and James E. Krier, *Environmental Law and Policy: Readings, Materials and Notes*, 2d ed., Indianapolis 1978.

Č. G. Veljanovski, "'Economic' Myths about Common Law Realities—Economic Efficiency and the Law of Torts," work. paper no. 5, Centre Socio-Legal Stud., Oxford Univ., Mar. 1979.

B. R. Weingast, M. E. McBride, and J. L. Conant, "Product Safety and Consumer Information: The Impact of Liability Assignment and Standards Regulation," work. paper no. 45, Center Stud. Amer. Bus., Washington Univ. (St. Louis), June 1979.

Vehicle and Nonvehicle Currencies in International Trade

By STEPHEN P. MAGEE AND RAMESH K. S. RAO*

> One has only to imagine what would happen to business calculations and plans if the number of ounces in a pound, or of inches in a foot, were...variable, and then to remember that, whereas these measures enter only into contracts concerning goods sold by weight or length, the monetary unit enters into every single economic contract of any kind whatever, to get an idea of the extent of the damage to economic efficiency for which a monetary system that is unreliable, or imperfectly understood, may be responsible. [Barbara Wooton, quoted in George Halm, p. 16]

There are two types of currencies in the pricing of internationally traded goods: vehicle and nonvehicle currencies. Nonvehicle currency pricing occurs when the currency of the exporter or the importer is used; vehicle currency pricing occurs when a third currency is used. Most internationally traded goods are not priced in dollars, but in nonvehicle currencies. We propose a division of nonvehicle currencies into three groups: major currencies, symmetric currencies, and minor currencies. For any pair of bilateral trading partners, one of the two currencies is defined as "major" if it is used as the dominant form of pricing for trade in both directions; the other currency is "minor." Bilateral trade between pairs of "symmetric currencies" occurs when one country's currency dominates trade in one direction while the other country's currency dominates trade in the other direction.

Using these definitions, we sketch the broad outlines of a theory of vehicle and nonvehicle currency trade which is consistent with the following stylized facts: for primary commodities, the dollar is an important vehicle currency in international trade; however, for most internationally traded manufactured goods, the dollar is not an important vehicle currency. In fact, for *U.S.* trade with about half of its developed country trading partners, the dollar is not even a major nonvehicle currency. Major-minor currency pricing relationships emerge between countries with strong and weak (low and high inflation) currencies, respectively, while symmetric currency pricing is the rule for countries with convertible currencies and equal inflation risk. For bilateral trade symmetrically invoiced in nonvehicle currencies, there is a tendency for pricing in the exporter's currency to predominate. This taxonomy is useful for analyzing such important problems as the *J*-curve effect, and how inflation risk influences the selection of a currency of invoice.

I. The Theory of Vehicle Currencies

In this section, consider the vehicle currency role of the dollar broadly to include interbank private and official financial transactions as well as private commodity trade. As international money, the dollar serves three important functions: a medium of exchange; a numeraire (also referred to as a unit of account or a standard of value); and a standard of deferred payment or store of value. Thus, we can identify the nine

*Professor of finance and assistant professor of finance, respectively, University of Texas-Austin. We are indebted to Gene Uselton and the Center for Education and Research in Free Enterprise, Texas A&M University, for research support; to Paul Krugman and especially Ronald McKinnon for helpful comments; to John Makin for generously sharing his data; and to Gunter Richter for research assistance. Any errors are ours.

different vehicle currency roles for the dollar classified by the type of transaction and the function of the currency. In rough summary, transactions costs explain the vehicle currency role of the dollar as a medium of exchange; information costs explain its vehicle role as a numeraire; and its intertemporal stability in real terms in the 1950's and 1960's explains its vehicle role as a standard of deferred payment.

Consider first why the dollar is the dominant medium of exchange in interbank clearing among convertible currencies. The most obvious theoretical explanation is based on transactions costs: the dollar will be a vehicle if it is cheaper to go from deutsche marks to dollars to British pounds than from deutsche marks directly to pounds. A less obvious result is Paul Krugman's finding that in general payments equilibrium, the dollar can still be a vehicle currency even when the indirect path (DM–\$–£) is more costly than the direct path (DM–£). The only necessary condition is a lower transactions cost for the dollar than for any other currency pair. While his results have not generalized beyond three currencies, Alexander Swoboda's data for the 1960's showed that this requirement held for the dollar vis-à-vis thirteen other countries.

Second, the dollar plays an important vehicle role as a numeraire or standard of value in certain product markets, but not in others. Work by Ronald McKinnon (1979), and Stephen Carse and Geoffrey Wood taken together indicates that international price quotations in homogeneous product markets, especially primary commodities, tend to be in vehicle currencies (dollars or sterling) while quotations in heterogeneous products tend to be in nonvehicle currencies. One reason for this result is that the United Kingdom and the United States are the locus of major world commodity exchanges: metals and tropical products are traded in London while grains, livestock, and metals are actively traded in New York and Chicago. More generally, homogeneity and international comparability of each primary product means that the demand for effective communication of price quotations leads to worldwide pricing in a common vehicle currency (this is a variation on McKinnon's 1969 information cost argument). With continuous changes in prices with the arrival of new information and through international arbitrage in primary commodities, it is more economical to transmit price change information in a single currency than through many. Since knowledge of a 1 percent rise in French francs is a perfect substitute for knowledge of a 1 percent rise in dollars, the world moves to a corner solution and uses a single currency for economy of communication. The same efficiency of communication argument applies to the pricing of liquid funds, that is, interest rates in the Euro-dollar market, as well as to the ease of official pegging of exchange rates by intervening in a single vehicle currency—normally the *U.S.* dollar.

In contrast, the geographical span of the market (that is, the area of a uniform price) for each heterogeneous manufactured product usually corresponds to the national economy in which it originates. These are "fix-price" goods (in the Hicksian sense) and McKinnon (1979, ch. 4) demonstrates that producers, whose factor costs are fixed in the domestic currency, reduce risk by keeping their invoice prices rigid also in domestic currency. For manufactured goods, rational price discrimination across currencies is frequent: the economic value of continuous price monitoring, as the exchange rate changes, is low. Thus, nonvehicle currency pricing, which is insensitive to "small" exchange-rate changes is the rule. One implication of these views is that the number of currencies used in international price quotations will be more or less equal to the number of separate national markets. For heterogeneous manufacturers, international pricing practices result in the natural emergence of rudimentary optimum currency areas, on a product-by-product basis.

Third, consider the vehicle role of the dollar as a standard of deferred payment. This applies to all dollar-denominated international trade which is contracted or invoiced for future payment: Euro-bonds and interest-bearing dollar assets held as re-

TABLE 1—THE CURRENCY OF INVOICING OF EXPORTS

	Page Data: Percentage of Country's Total Exports Invoiced in:			Makin Data: Country's Exports to United States, Percentage Invoiced in:[a]		Percent in Dollars		Appreciation of the Country's Currency, 1970–75 ($/currency)[b]
	Exporter's Currency (1)	Importer's Currency (2)	U.S. Dollars (3)	Exporter's Currency (4)	U.S Dollars (5)	1972 (6)	1975 (7)	(8)
Belgium	48		11	22	78	98	57	35
Canada				13	87	93	81	3
France	68	24	9	48	52	60	45	29
Italy				32	68	90	46	−4
Japan				6	94	99	89	21
Netherlands	50		13	20	81	83	68	43
Sweden	66	25	12	58	27	4	–	25
Switzerland				72	28	45	10	67
United Kingdom	73	15	18	51	44	36	52	8
West Germany	87		5	64	36	38	34	48
Simple Average	65		11	39	60			

[a]Average of the 1972 and 1975 percentages.
[b]Shown in percent.

serves by central banks. The financial patterns in international trade just discussed carry over to this case as well; namely, homogeneous products tend to be contracted for in dollars. Moreover, the use of the dollar as a convenient vehicle for official intervention leads to a derived demand by central banks throughout the world to hold it as a reserve asset (see McKinnon, 1980). We defer until the next section a discussion of the Bilson-Magee hypothesis which asserts that the demand for a contracting currency, whether vehicle or nonvehicle (as well as the demand for nominal bonds and official reserves), is positively related to its intertemporal stability in real terms.

II. The Theory of Nonvehicle Currencies

The theory of nonvehicle currencies —major, minor, or symmetric—applies to all trade in which the goods are priced in the currency of one of the two trading partners. We address four issues: the empirical evidence; the Bilson-Magee theory of major and minor nonvehicle currencies; the *J*-curve and foreign exchange market implications of Grassman-type symmetries; and two theories of symmetric currency pricing.

Consider the still rather scarce empirical evidence. It appears that most *U.S.* exports are denominated in *U.S.* dollars. John Makin's data in column (5) of Table 1 suggests that the dollar is a *major* nonvehicle currency, dominating both *U.S.* export and import pricing, vis-à-vis Belgium, Canada, France, Italy, Japan, and the Netherlands; these countries are minor nonvehicle currencies vis-à-vis the United States. The dollar is a *symmetric* nonvehicle currency, with *U.S.* exports primarily in dollars and *U.S.* imports primarily in the exporter's currency, for *U.S.* trade with Sweden, Switzerland, the United Kingdom, and Germany. We know of no cases in which the dollar is a minor nonvehicle currency in its bilateral trade.

Among developed countries other than the United States, Sven Grassman was the first to discover that such trade is largely symmetric. His study of Sweden indicated that two-thirds of trade is denominated in the exporter's currency; one-fourth is denominated in the importer's currency and a trivial proportion in vehicle currencies, primarily dollars. The data by S. A. B. Page in Table 1 shows that the rule appears to hold for five other Western European countries. For trade between the developed and developing countries, very little pricing is done in *LDC* currencies. In some cases involving primary products, dollars or sterling are used; but in most others, the

currency of the developed country trading partner predominates. If this is the case, *DC-LDC* trade is best described as a major-minor nonvehicle currency relationship.

Notice that for all world trade, the length of the lag between order and payment affects the foreign exchange risk. Magee (1974) reports that these lags vary across countries (for *U.S.* imports, the order to delivery lags equal 4 months from Germany and 4.5 months from Japan), but they vary even more across industries (for *U.S.* imports from Germany, the lags are 1 month for autos, 5 months for steel and 9 months for machine tools). Thus, nonvehicle currency pricing is related to the value of the currencies involved as a standard of deferred payment.

Next consider the Bilson-Magee hypothesis (reported in Rao and Magee, 1980) that all traders (including importers) in high inflation-risk countries will prefer to price in the stronger (low inflation or appreciating) currency of the trading partner. The hypothesis explains why most bilateral trade between *LDC*s and *DC*s is invoiced in the *DC*'s currency, and why external borrowing by *LDC* nationals occurs only in hard currencies. A key to the concept is understanding that international trade contracts, like nominal bonds, are subject to inflation risk or the risk of exchange controls. If a Brazilian importer agrees to purchase goods in the United States for future delivery and payment one year from now, would he want to pay in the strong currency (dollars) or in his weak currency (cruzeiros)? Assume that the expected inflation rate in Brazil over the next year is 60 percent and that by the Fisher effect, this increase is incorporated into the contract price. If the realized inflation rate equals 60 percent, there are no surprises. But, with high inflation risk, the realized real value of a cruzeiro contract can have considerable variance. If the inflation rate turns out to be 90 percent, the importer gains from cruzeiro pricing since his real payment is only .84 $(=1.6/1.9)$ of today's price of the good. But, if the realized inflation rate turns out to be only 30 percent, his real payment is 1.23 $(1.6/1.3)$ times today's price of the good (as well as 1.23 times his expectation). If, however, the realized monetary shocks that cause the unexpected variations in inflation also cause parallel changes in foreign exchange rates (i.e., assuming full purchasing power parity (*PPP*)), the variance in the real cost of a foreign currency (dollar) denominated contract will be zero and it will be preferred. The reason is that with 90 percent inflation, the thirty percentage points of unexpected inflation will be offset by thirty percentage points of unexpected cruzeiro depreciation. The same argument holds with signs reversed for a realized inflation of only 30 percent. Because *PPP* does not hold exactly in practice, the whole matter boils down to the empirical question of whether the deviations from purchasing power parity are more or less than domestic inflation risk. Notice that, as stated, the Bilson-Magee result implies a corner solution: that is, foreign currency denominated contracts will dominate even the inflation-risk-adjusted *LDC* contract prices. Evidence from the German hyperinflation of the 1920's substantiates this effect: Germans quoted contract prices to each other in dollars, while continuing to use marks as the medium of exchange.

Although this proposition is obvious for developing countries (who have both weak currencies and exchange controls) vis-à-vis the developed countries, does it apply to trade *among* developed countries? The answer is yes: major nonvehicle currencies can descend to symmetric nonvehicle currencies, and the latter can drop to minor nonvehicle currencies, with increases in home inflation risk. Higher relative *U.S.* inflation risk in the 1970's has decreased dollar invoicing and contracting over the longer term. From 1972 to 1975, Table 1 shows a decline in the proportion of *U.S.* imports contracted in dollars with every country except the United Kingdom. Further, these decreases were correlated with the relative excess of *U.S.* inflation as measured by the appreciation of the country's currency shown in column (8): the simple correlation is nearly $-.7$ when Canada and Italy are dropped (the Canadian dollar is too near a substitute to the *U.S.* dollar and Italy directly restricted

dollar contracting by controls in the 1970's). Although the calculation is very rough, the regression coefficient in this relationship suggests that each percentage point rise in the dollar cost of a foreign currency (depreciation of the dollar) with a trading partner leads to a 1.4 percentage point decline in the use of the dollar as the contracting currency. For example, in German-American trade the use of the deutsche mark has increased sharply. In summary, the dollar is slipping in its nonvehicle currency role.

How does the selection of a currency of invoice influence the trade balance following an unanticipated exchange in the foreign exchange rate? Magee (1973, 1974) suggested that a Grassman pattern of predominantly exporter (symmetric) currency contracting would generate a *J*-curve for a country's trade balance, which declines initially and then later improves following an unanticipated devaluation. To illustrate the contracting effect, consider the extreme case in which all trade is invoiced in the exporter's currency; for example, all *U.S.* exports to Germany are in dollars and all *U.S.* imports from Germany are in *DM*. If the dollar depreciated vis-à-vis the *DM*, the value of outstanding *U.S.* exports would be unaffected since they are invoiced in dollars, but all outstanding *U.S.* import contracts would have a higher dollar cost than originally expected because of the new higher dollar cost of the *DM*. Thus, the *U.S.* trade balance in dollars would initially decline. Carse and Wood have extended this argument: the *J*-curve phenomenon will have destabilizing effects on foreign exchange markets. However, with efficient markets, their result implies greater variance of rates only if the phenomenon is not understood.

Finally, consider two theories which have emerged to explain symmetric nonvehicle currency pricing. In the first, McKinnon (1979), has developed arguments explaining the tendency toward exporter currency pricing discovered by Grassman. McKinnon hypothesizes that exporters of manufactures will have a greater interest in invoicing in their home currency. Importers can raise the domestic price of the imported good if an unexpected devaluation occurs; whereas an exporter cannot cut the price of contractually set wage payments and interest payments in response to a revaluation induced decline in the home currency price of his exports. In effect the variance in an exporter's output price contributes more to the variance of a firm's cash flow than equivalent variance in an importer's purchasing costs. This differential preference by exporters for contracting in their home currency shifts exchange risk to importers. McKinnon argues that this shift is effected by exporters offering importers the quid pro quo of open-account trade credit: thus, importers take the foreign exchange risk, but pay the exporter whenever they want. This payments flexibility makes it much easier for the importer to reduce exchange risk by hedging. While there is empirical support for McKinnon's two hypotheses, they are vulnerable to efficient market arguments that foreign exchange risk is frequently "unsystematic" and could be eliminated by simple portfolio diversification.

A second theory of symmetric nonvehicle currency pricing, developed in our 1980 paper, suggests that there is no reason to expect either predominantly exporter or importer currency pricing between countries with equal inflation risk. We argue that Grassman's two-thirds rule might be spurious and present data from ten countries indicating that the proportion of trade invoiced in the exporters' currency is not significantly different from one-half, that is, from a random distribution. We then derive an irrelevance proposition which states that even if the two traditional hedging mechanisms are not available (forward markets or international lending and borrowing), the currency of denomination of international trade contracts may still be irrelevant to a representative exporter or importer. However, one assumption is quite restrictive: the result requires that all traders in both countries are nearly identical in their degrees of risk aversion. It further assumes that there is no asymmetry in production costs on cash flows, as described above,

between exporters and importers. Then, consider a *U.S.* exporter and a *U.K.* importer negotiating a contract for delivery and payment in one year. Assume that the forward exchange rate is 1; that the dollar contract price is \$100, and the sterling contract price is £105. Thus, if the *U.S.* exporter contracts in dollars he receives \$100 with certainty; if he contracts in sterling, he will receive a payment whose expected value today is \$105. Similarly, the *U.K.* importer faces a certain payment of £105 if the contract is in sterling or an uncertain payment whose expected value is £100 if the contract is in dollars. Assume that the degree of risk aversion is such that at these prices both *U.S* exporters and *U.K.* importers are just compensated for the risks of contracting in foreign currency, that is, \$100 is the certainty equivalent of a sterling contract for *U.S.* exporters and £105 is the certainty equivalent of a dollar contract for *U.K.* importers. In this case, \$100 is the equilibrium dollar contract price and £105 is the equilibrium sterling price. Since all parties in the United States and all parties in the United Kingdom are identical, parties on both sides of the Atlantic are indifferent between dollar and sterling contracts.

But once one introduces cost asymmetries between importers and exporters, different degrees of inflation risk associated with various currencies, and distinctions between homogeneous primary commodities and manufactured goods, then our classification into major, minor, symmetric, and vehicle currencies becomes useful.

REFERENCES

S. Carse and G. E. Wood, "Currency of Invoicing and Forward Covering: Risk-Reducing Techniques in British Foreign Trade," in John P. Martin and Alasdair Smith, eds., *Trade and Payments Adjustment under Flexible Exchange Rates*, London 1979, pp. 174–96.

S. Grassman, "A Fundamental Symmetry in International Payment Patterns," *J. Int. Econ.*, May 1973, *3*, 105–16.

George N. Halm, *Monetary Theory*, Philadelphia 1942.

P. Krugman, "Vehicle Currencies in the Structure of International Exchange," mimeo., Yale Univ. 1979.

S. P. Magee, "Currency Contracts, Pass-Through and Devaluation," *Brookings Papers*, Washington 1973, *1*, 303–23.

______, "U.S. Import Prices in the Currency Contract Period," *Brookings Papers*, Washington 1974, *1*, 117–64.

______, "Contracting and Spurious Deviations from Purchasing-Power-Parity," in Jacob A. Frenkel and Harry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*, Reading 1978, 67–74.

Ronald McKinnon, "Private and Official International Money: The Case of the Dollar," *Essays in International Finance*, Princeton Univ., No. 74, Apr. 1969.

______, *Money in International Exchange*, New York 1979.

______, "Dollar Stabilization and American Monetary Policy," *Amer. Econ. Rev. Proc.*, May 1980, *70*, 382–87.

S. A. B. Page, "The Currency of Invoicing in Merchandise Trade," *Nat. Inst. Econ. Rev.*, No. 81, Aug. 1977, 77–81.

R. K. S. Rao, and S. P. Magee, "The Currency of Denomination of International Trade Contracts," in Richard M. Levich and Clas G. Wihlborg, eds., *Exchange Risk and Exposure: Current Developments in International Financial Management*, Lexington 1980.

A. Swoboda, "Vehicle Currencies in the Foreign Exchange Market: The Case of the Dollar," in Robert Z. Aliber, ed., *The International Market for Foreign Exchange*, New York 1969, 30–40.

The Efficiency of Foreign Exchange Markets and Measures of Turbulence

By JACOB A. FRENKEL AND MICHAEL L. MUSSA*

Since the move to generalized floating in 1973, exchange rates between major currencies have displayed large fluctuations. This turbulence of foreign exchange rates is an important concern of government policy and its explanation is a challenge for theories of foreign exchange market behavior. In Section I of this paper, we document the extent of turbulence in foreign exchange markets by examining (i) the magnitude of short-run variations in exchange rates relative to other measures of economic variability; (ii) the degree of divergence between actual and expected changes in exchange rates; and (iii) the extent to which exchange-rate movements have diverged from movements of relative national price levels. In Section II, we provide a general explanation of this turbulence in terms of the modern "asset market theory" to exchange-rate determination. This theory emphasizes that exchange rates, like the prices of other assets determined in organized markets, are strongly influenced by the market's expectation of future events. In this context, we also discuss the narrower technical question of "foreign exchange market efficiency." Finally, in Section III, we address the question of whether turbulence in the foreign exchange markets has been "excessive" and what policy measures can (or should) be taken to reduce it.

I. Measures of Turbulence

Perhaps the simplest measure of turbulence in foreign exchange markets is the average percentage changes in exchange rates over some interval of time. For example, Table 1 reports the average absolute monthly percentage change for three major exchange rates (the dollar/pound, the dollar/French franc and the dollar /deutsche mark (*DM*)) for the period June 1973 to February 1979. In all cases, the average absolute change exceeded 2 percent per month. In comparison the average absolute monthly percentage change for wholesale and consumer price indices and for ratios of national price levels were only about half that of the exchange rates, and the differences are even more striking for the detrended series.

Another important dimension of exchange-rate turbulence is the predictability of exchange-rate changes. If exchange-rate changes were largely predictable, they would give rise to little risk. The facts indicate, however, that exchange-rate changes are largely unpredictable. If we regard the forward premium on foreign exchange as a measure of the market's prediction of the change in the exchange rate, then predicted changes in the exchange rate account for a very small fraction of actual changes. This general phenomenon is illustrated for the case of the dollar/*DM* exchange rate in Figure 1.

If exchange rates moved in accord with relative national price levels, as suggested by a simple version of the purchasing power parity theory, the turbulence of exchange rates would be regarded as a manifestation of the forces underlying the turbulence of national inflation rates. Moreover, in this circumstance, turbulence of exchange rates would probably not be regarded as an addi-

*University of Chicago and the National Bureau of Economic Research, Inc; and the University of Chicago, respectively. We are indebted to Lauren Feinstone for efficient research assistance and to members of the International Economics Workshop at the University of Chicago for helpful comments. Frenkel acknowledges research support by the National Science Foundation, grant SOC 78-14480. This research is part of the *NBER*'s Program in International Studies. The views expressed are those of the authors and not necessarily those of the *NBER*.

TABLE 1—MEAN ABSOLUTE PERCENTAGE CHANGES IN PRICES AND EXCHANGE RATES
Monthly Data: June 1973–February 1979

Country	*WPI*	*COL*	Stock Market	Exchange Rate Against the Dollar	COL/COL_{US}
United States	.009	.007	.038	–	–
United Kingdom	.014	.012	.066	.020	.007
France	.011	.008	.054	.020	.003
Germany	.004	.004	.031	.024	.004

Note: All variables represent the absolute values of monthly percentage changes in the data. *WPI* denotes the Wholesale Price Index and *COL* denotes the Cost-of-Living Index. Data on prices and exchange rates are from the IMF tape (May 1979 version). The stock market indices are from *Capital International Perspective*, monthly issues.

tional source of social cost. However, as illustrated in Figure 1, short-run changes in exchange rates bear little relationship to short-run differentials in national inflation rates, particularly as measured by consumer price indices. Further, as illustrated in Figure 2 for the case of Germany and the United States, changes in exchange rates over longer periods of time have frequently been associated with large cumulative divergences from relative purchasing power parities.[1]

In summary, turbulence of exchange rates is indicated by the large and unpredictable fluctuations that do not conform closely to movements in relative national price levels. The question remains—what is the social cost of this turbulence? It appears to have significant social cost. It generates capital gains and losses for holders of assets denominated in different national monies, and presumably induces wealth holders to alter their behavior and expend resources in order to reduce risk. By interfering with the efficiency of the price system in guiding resource allocation, such turbulence may also lead to economically inappropriate patterns of production, consumption, and trade. The issue of social cost, however, involves the more subtle question of whether turbulence has been "excessive." We will turn to this question after we have discussed the analytical perspective from which exchange-rate turbulence may best be understood.

II. Exchange Rates as Asset Prices

The "asset market theory" of exchange rates recognizes that exchange rates are relative prices of assets determined in organized markets where prices can be adjusted on a moment-to-moment basis to whatever "the market" regards as the currently appropriate price. In this fundamental respect, exchange rates are similar to stock prices, long-term bond prices, and the prices of commodities traded on organized exchanges.[2]

The most important implications of the asset market theory of the exchange rate can be exposited in a simple model. Suppose that the logarithm of the spot exchange rate on day t, $s(t)$, is determined by

$$(1)\quad s(\mathrm{t})=z(\mathrm{t})+b\cdot E[s(\mathrm{t}+1)-s(\mathrm{t});\mathrm{t}]$$

where $E[s(\mathrm{t}+1)-s(\mathrm{t});\mathrm{t}]$ denotes the expected percentage change in the exchange rate between t and t+1, based on information available at t, and where $z(\mathrm{t})$ represents the ordinary factors of supply and demand that affect the exchange rate on day t

[1]Related measures of turbulence are discussed by Robert Aliber and Ronald McKinnon. For an assessment of the evidence on purchasing power parities, see the various contributions to the symposium published in the *Journal of International Economics*, May 1978.

[2]For theoretical developments and applications of the asset market theory of exchange-rate determination, see Rudiger Dornbusch (1976a), Frenkel (1976), Pentti Kouri, Mussa (1976a), and William Branson. For collections of articles representing the asset market theory, see the *Scandinavian Journal of Economics*, No. 2, (May 1976), and Frenkel and Harry Johnson.

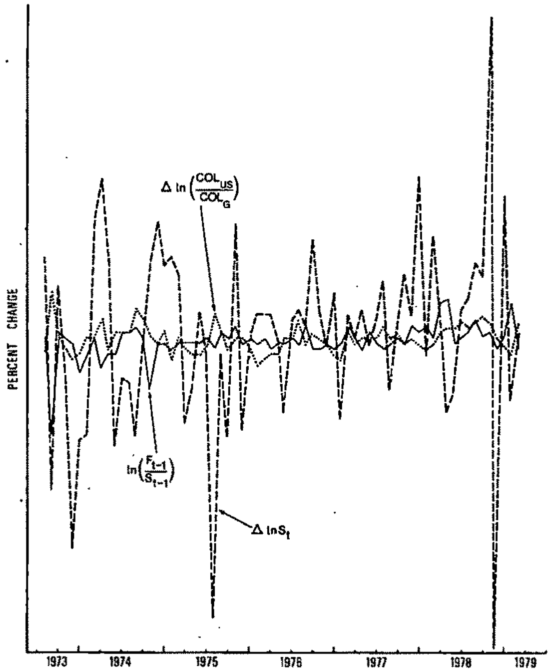


FIGURE 1. Monthly Percentage Changes of the U.S./German Consumer Price Indices, of the $/*DM* Exchange Rate and the Monthly Forward Premium; July 1973–Feb. 1979

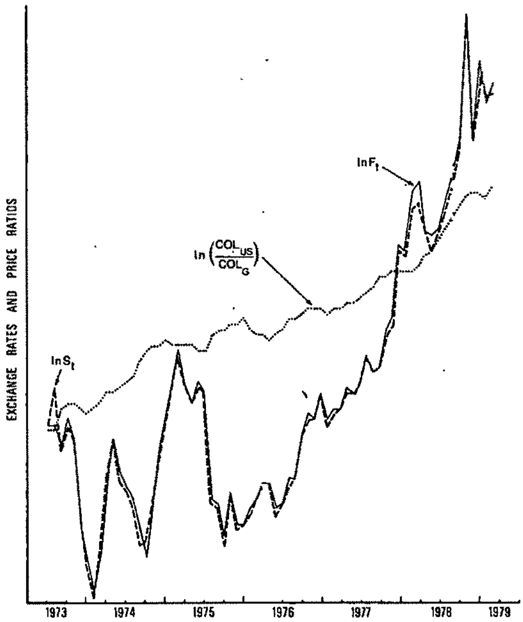


FIGURE 2. Monthly Observations of the Dollar/*DM* Spot and Forward Exchange Rates and the Ratio of the U.S./German Cost of Living Indices (scaled to equal the spot exchange rate at the initial month): June 1973–Feb. 1979

(among which are domestic and foreign money supplies, incomes, output levels, and the like).[3] Assuming that expectations are "rational" in that equation (1) applies to expectations of future exchange rates, it follows, by forward iteration, that[4]

$$(2) \quad E(s(t+j);t)=(1/(1+b)) \sum_{k=0}^{\infty} (b/(1+b))^k \cdot E(z(t+j+k);t)$$

Thus, the current exchange rate ($j=0$) and current expectations of future exchange rates ($j>0$) are linked because both depend on expectations concerning the future zs. The presumption is that this linkage should be quite strong at least for the exchange rates expected in the near future. The reasoning is straightforward. If the expected change in the exchange rate over a single day is 1 percent, the expected rate of return from shifting wealth from one currency of denomination to another is extremely high. The flow of funds induced by such a high expected rate of return should swamp the ordinary factors of supply and demand represented by $z(t)$. Hence, the current exchange rate, $s(t)=E(s(t);t)$, should be closely linked to the current expectation of tomorrow's exchange rate, $E(s(t+1);t)$, which in turn should be closely linked to the exchange rate expected for the following day, $E(s(t+2);t)$; and so on.[5] Further, using

[3]Equation (1) is a reduced form that can be derived from a variety of models of exchange rate-determination. The various models may differ in their emphasis of the determinants of $z(t)$ but they all share a similar reduced form.

[4]A result of this general form is derived in Mussa (1976a, 1977a). The unique role played by expectations is also emphasized by Stanley Black, Dornbusch (1976b), and John Bilson.

[5]If the time period is literally a day, then for any reasonable model the coefficient b should be numerically very large. Hence, the weight given to all expected future zs in determining the current exchange rate, $b/(1+b)$, will also be relatively large.

equation (2) we may determine the expected change in the exchange rate, conditional on information available at time t. The actual change is this expected change plus the (necessarily unpredictable) unexpected change due to "new information" that alters expectations concerning the zs.

The evidence from the behavior of spot and forward rates, illustrated in Figure 1, indicates that the bulk of exchange-rate changes appear to be due to new information. The close correlation between movements of spot and forward rates illustrated in Figure 2 indicates that this new information appears to alter views concerning current and expected future exchange rates by approximately the same amount. This close correlation between spot and forward rates is also evidence of the close link between current and expected future exchange rates. Moreover, while the dominance of unexpected changes and the close correlation of spot and forward rates are not theoretical necessities, it is noteworthy that these characteristics are shared by the prices of many assets and commodities traded in organized markets.[6]

The principle that exchange rates change in response to new information is relevant for understanding the relationship between the trade balance and exchange-rate changes. This principle suggests that a deficit in the trade balance may or may not be associated with depreciation, depending on whether this deficit was unexpected or expected. More generally, since exchange rates already reflect expectations of future deficits or surpluses, changes in exchange rates will depend largely on the unexpected component of the trade balance and on the extent to which the new information embodied in this unexpected component affects expectations concerning future trade imbalances.

Within the broad perspective of foreign exchange markets as asset markets, there is the narrower technical question of "market efficiency." For an asset market to be "efficient," prices must appropriately reflect all available information and thus it should be impossible to make extraordinary profits by exploiting generally available information. Tests of efficiency that have focused on the possibility of making extraordinary profits by exploiting covered interest arbitrage opportunities have generally concluded that foreign exchange markets are efficient, at least when account is taken of the costs of making transactions in various markets. Other tests of foreign exchange market efficiency have focused on (i) the statistical properties of forward rates as predictors of future spot rates, (ii) the time-series properties of exchange rates and of deviations of exchange rates from past forward rates, (iii) the ability to improve on market forecasts of future exchange rates by using past spot and forward exchange rates and other publicly available information, and (iv) the capacity to make extraordinary profits by employing various trading rules. Different tests applied to different exchange rates in different time periods have not reached a unanimous consensus concerning the hypothesis of foreign exchange market efficiency.[7]

Whatever the outcome of the debate on the technical question of market efficiency, the general view of exchange rates as asset prices is vital in explaining the turbulence of exchange rates documented in Section I. This perspective implies that exchange rates will not adjust slowly and smoothly, but like other asset prices, will display random fluctuations in response to new information that is continually being received by the market. While it is exceedingly difficult to determine the precise new information which caused particular changes in individual exchange rates, the turbulent events since 1973 are consistent with substantial

[6]The hypothesis that the predominant cause of exchange-rate movements is "news" which could not have been anticipated has been put forward by Mussa (1976b, 1977b, 1979) and Dornbusch (1978). For an empirical test of the hypothesis, see Frenkel (1980).

[7]Tests of efficiency of interest arbitrage are reported by Frenkel and Richard Levich. Tests of efficiency of the foreign exchange market for various periods are reported by Frenkel (1977, 1978, 1980, forthcoming), Frenkel and Kenneth Clements, Paul Krugman, Maurice Obstfeld, Craig Hakkio and by Lars Hansen and Robert Hodrick. For surveys, see Levich and Steven Kolhagen.

volatility of exchange rates. During this period, many countries experienced the widest variations of output, employment, and inflation rates of the post-World War II era. In virtually all countries, there have been wide swings in government policy. The world economy has had to adjust to the real and financial consequences of the increase in the price of oil. Exchange rates, like other asset prices, have responded not only to events but also to their effect on expectations concerning future events. In short, there is no great mystery why exchange rates have shown a high degree of turbulence during recent years.

Finally, divergences of exchange rates from relative purchasing power parities are not necessarily indicative of any peculiarity in the behavior of exchange rates. Time-series analysis reveals that monthly changes in exchange rates, like changes in many other asset prices, show little or no serial correlation. Measured national price levels, however, do *not* behave like exchange rates, but instead exhibit a degree of serial correlation. Recent macro-economic theorizing has rationalized this "stickiness" in terms of nominal contracts, costs of price adjustments, as well as confusion between relative and absolute prices and between permanent and transitory changes. The difference between the time-series properties of exchange rates and national price levels explains why there is little correlation between the essentially random month-to-month changes in exchange rates and slowly moving (serially correlated) inflation rate differentials. If two or three monthly changes in an exchange rate happen to be large and in the same direction, then given the stickiness of national price levels it is likely that there will be a large cumulative divergence from relative purchasing power parity.

In summary, the asset market theory of exchange rates explains the three key facts concerning exchange-rate turbulence that were examined in Section I.

III. Is Turbulence Excessive?

The asset market theory does not provide definitive answers to two other important questions: Have exchange rates fluctuated "excessively" since 1973? And could government policy be better managed to reduce the extent and cost of exchange-rate turbulence?

Concerning the first of these questions, four points should be made. First, as previously noted, the period since 1973 witnessed great turbulence in the world economy and great uncertainty about the future course of economic events. In this environment, all asset prices, not only exchange rates, have shown large fluctuations. Indeed, while exchange-rate changes have been large in comparison with changes in national price levels, they have been considerably smaller than changes in the prices of other assets like gold, many other commodities, and common stocks (see Table 1).

Second, while it is difficult to identify and measure the cost of exchange-rate turbulence, it is clear that dire predictions of the collapse of world trade and the disintegration of world financial markets following the move to floating exchange rates have proved exaggerated. World trade has continued to expand, and capital markets have shown considerable resilience in dealing with the strains created by the oil crisis. It is doubtful that the system of pegged rates could have survived in that environment without severe limits on trade and capital movements being imposed by many countries.

Third, changes in real economic conditions requiring adjustments in the *relative* prices of different national outputs occur continuously. Under the system of pegged rates, relative price adjustments are achieved through the slow changes of national price levels and through occasional changes of parity. Under floating rates, adjustments in the relative price of different national outputs occur rapidly and in anticipation of changes in economic conditions rather than after the need for adjustment has become apparent. In the absence of an explicit specification of relative costs, there is no general presumption that slow adjustment of relative prices is preferable to rapid adjustment, or that price adjustments should not occur in anticipation of events requiring

such adjustments. Indeed, since prices play an important role in communicating the appropriate manner to deploy resources, rapid and anticipatory adjustment of relative prices may contribute to the efficiency of the economic system. Given the apparent stickiness of national price levels, the flexibility of relative prices of national outputs provided by floating exchange rates may contribute to the efficient functioning of the economic system.

Fourth, part of the variability of exchange rates and associated variability of relative prices of national outputs may represent an "overshooting" response to purely monetary disturbances. This suggests that exchange-rate changes induced by purely monetary disturbances could be a source of costly real disturbances. However, since overshooting can only be defined with respect to the equilibrium exchange rate, it is essential to have a well-defined measure of the equilibrium exchange rate in order to judge the empirical relevance of overshooting. At present, there is no reliable evidence that exchange-rate changes are dominated by overshooting, rather than by adjustments to new information about equilibrium exchange rates. Moreover, even if overshooting were empirically important, it would not automatically follow that policies should be directed toward eliminating it. If overshooting reflects the failure of domestic price levels to adjust to changes in current and expected economic conditions, it may be desirable to allow excessive adjustment in some other prices.

The final question we wish to address is how government policy can be managed to reduce costly and undesirable turbulence of foreign exchange rates. Concerning this question, we shall make four points. First, exchange rates are macro-economic variables that reflect the (actual and anticipated) behavior of both absolute and relative prices in different countries, and are influenced by the whole array of (actual and expected) government policies, especially policies that affect the demand and supply of different national monies. Exchange rates, however, are not tools of policy that may be manipulated independently of other policy tools to achieve desired objectives. Hence, it makes no sense to talk about "exchange-rate policy" as an independent instrument of government policy in isolation from other elements of macro-economic policy.

Second, while as a technical matter, government policy can reduce exchange-rate fluctuations, even to the extent of pegging an exchange rate, it may not be assumed that such policies will automatically eliminate the disturbances that are presently reflected in the turbulence of exchange rates. Such policies may only transfer the effect of disturbances from the foreign exchange market to somewhere else in the economic system. There is no presumption that transferring disturbances will reduce their overall impact and lower their social cost. Indeed, since the foreign exchange market is a market in which risk can easily be bought and sold, it may be sensible to concentrate disturbances in this market, rather than transfer them to other markets, such as labor markets, where they cannot be dealt with in as efficient a manner.

Third, one exception to this general principle is that government policy can offset disturbances arising from shifts in demands to hold different national monies. The correct policy is to accommodate such demand shifts by offsetting supply shifts, thereby eliminating the need for costly adjustments of exchange rates and national price levels. The difficulty with implementing this policy is in identifying when a shift in money demand has occurred and in adjusting the asset side of the balance sheets of national banking systems.

Fourth, another way in which government policy can make a positive contribution to reducing costly and unnecessary turbulence of foreign exchange rates is by reducing high and variable rates of monetary expansion which, for example, result from misguided attempts to stabilize nominal interest rates. This is especially important because, as follows from the asset market theory, exchange rates are affected not only by current policy actions, but also by current expectations of future policy. If expectations of future policy are highly sensitive to cur-

rent policy, then instability of policy can have a magnified effect on exchange rates and on the relative prices of different national outputs, thereby generating significant social costs. If, as we believe, the instability and unpredictability of policy, particularly monetary policy, has contributed significantly to the turbulence of exchange rates since 1973, then the turbulence and its associated cost can be reduced by adopting more stable and predictable patterns of government policy. This is particularly important for the United States in view of its size and in view of the special role of the dollar in the international monetary system. The *U.S.* dollar has performed the role of "world money" in serving the functions of invoice currency, unit of account, intervention currency, store of value, international reserve asset, and the like. The principal contribution that *U.S.* economic policy can make in reducing exchange-rate turbulence and achieving greater economic stability, is by reducing the level and variability of *U.S.* inflation and simultaneously reducing uncertainty about *U.S.* economic policy. This should increase the efficiency and enhance the role of the dollar as world money.

REFERENCES

R. Z. Aliber, "The Firm Under Fixed and Flexible Exchange Rates," *Scand. J. Econ.*, No. 2, 1976, *78*, 309–22.

J. F. O. Bilson, "Rational Expectations and the Exchange Rate," in Jacob A. Frenkel and Harry G. Johnson, eds., *The Economics of Exchange Rates: Selected Studies*, Reading 1978.

S. W. Black, "International Money Markets and Flexible Exchange Rates," *Essays in International Finance*, Princeton Univ., No. 32, Mar. 1973.

W. H. Branson, "Asset Markets and Relative Prices in Exchange Rate Determination," *Sozialwissenschaftliche Annalen*, 1977, 69–89.

R. Dornbusch, (1976a) "The Theory of Flexible Exchange Rate Regimes and Macroeconomic Policy," *Scan. J. Econ.*, No. 2, 1976, *78*, 255–75.

______, (1976b) "Expectations and Exchange Rate Dynamics," *J. Polit. Econ.*, Dec. 1976, *84*, 1161–76.

______, "Monetary Policy Under Exchange Rate Flexibility," *Managed Exchange Rate Flexibility: The Recent Experience*, Fed. Reserve Bank Boston Conference Series, No. 20, 1978.

Jacob A. Frenkel, "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence, "*Scand. J. Econ.*, No. 2, 1976, *78*, 200–24.

______, "The Forward Exchange Rate, Expectations and the Demand for Money: The German Hyperinflation," *Amer. Econ. Rev.*, Sept. 1977, *67*, 653–70.

______, "Purchasing Power Parity: Doctrinal Perspective and Evidence from the 1920's," *J. Int. Econ.*, May 1978, *8*, 169–91.

______, "Flexible Exchange Rates in the 1970's," in *Stabilization Policies: Lessons from the 1970's and Implications for the 1980's*, St. Louis 1980.

______, "The Forward Premium on Foreign Exchange and Currency Depreciation During the German Hyperinflation," *Amer. Econ. Rev.*, forthcoming.

______ **and K. C. Clements**, "Exchange Rates in the 1920's: A Monetary Approach," in M. June Flanders and Assaf Razin, eds., *Development in an Inflationary World*, New York 1980.

______ **and Harry G. Johnson**, *The Economics of Exchange Rates: Selected Studies*, Reading 1978.

______ **and R. M. Levich**, "Transactions Costs and Interest Arbitrage: Tranquil Versus Turbulent Periods," *J. Polit. Econ.*, Dec. 1977, *85*, 1209–26.

L. P. Hansen and R. J. Hodrick, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *J. Polit. Econ.*, Aug. 1980.

C. S. Hakkio, "Expectations and the Foreign Exchange Market," unpublished doctoral dissertation, Univ. Chicago 1979.

Steven W. Kolhagen, *The Behavior of Foreign Exchange Markets—A Critical Survey of the Empirical Literature*, Monog. Series Finance Econ., No. 1978-3, New York Univ. 1978.

P. J. K. Kouri, "The Exchange Rate and the Balance of Payments in the Short Run and in the Long Run: A Monetary Approach," *Scand. J. Econ.*, No. 2, 1976, *78*, 280–304.

P. Krugman, "The Efficiency of the Forward Exchange Market: Evidence from the Twenties and the Seventies," unpublished manuscript, Yale Univ. 1977.

R. M. Levich, "The Efficiency of Markets for Foreign Exchange," in Jacob A. Frenkel and Rudiger Dornbusch, eds., *International Economic Policy: Theory and Evidence*, Baltimore 1979.

R. I. McKinnon, "Floating Exchange Rates 1973–74: The Emperor's New Clothes," in Karl Brunner and Allan Meltzer, eds., *Institutional Arrangements and the Inflation Problem*, Vol. 3, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 79–114.

M. L. Mussa, (1976a) "The Exchange Rate, the Balance of Payments and Monetary and Fiscal Policy Under a Regime of Controlled Floating," *Scand. J. Econ.*, May 1976, *78*, 229–48.

______, (1976b) "Our Recent Experience with Fixed and Flexible Exchange Rates: A Comment," in Karl Brunner and Allan Meltzer, eds., *Institutional Arrangements and the Inflation Problem*, Vol. 3, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1976, 123–41.

______, (1977a) "A Dynamic Theory of Foreign Exchange," in Michael Artis and A. Robert Nobay, eds., *Studies in Modern Economic Analysis*, Oxford 1977.

______, (1977b) "Exchange Rate Uncertainty: Causes, Consequences and Policy Implications," unpublished manuscript, Univ. Chicago 1977.

______, "Empirical Regularities in the Behavior of Exchange Rates and Theories of the Foreign Exchange Market," in Karl Brunner and Allan Meltzer, eds., *Policies for Employment, Prices, and Exchange Rates*, Vol. 11, Carnegie-Rochester Conferences on Public Policy, *J. Monet. Econ.*, Suppl. 1979, 9–57.

M. Obstfeld, "Expectations and Efficiency in the Foreign Exchange Market," unpublished manuscript, Mass. Inst. Technology 1978.

Dollar Stabilization and American Monetary Policy

By RONALD I. MCKINNON*

In the 1950's and 1960's a strong dollar standard existed; under fixed exchange rates the monetary policies of most nations were governed by the stable monetary policy of the United States. In the 1970's, the dollar standard was reduced to a weak form where industrial (but not less-developed) countries float their exchange rates to secure more domestic monetary independence, but the financial processes underlying international trade itself remain dollar based.

What are the elements of the weak dollar standard necessary for world trade to remain largely monetized and multilateral? How can nations avoid a relapse to the bilateralism and barter that characterized the 1930's? I shall consider first, necessary restraints on foreign exchange intervention and reserve holdings by central banks of industrial economies other than the United States; and second, the proper American monetary policy for reconciling the weakened international role of the dollar with domestic monetary stability.

I. Necessary Restraints on Central Banks from Industrial Economies

Let us divide countries into two groups: Group I is composed of industrial economies, which issue convertible currencies usable internationally by third countries and by each other. Merchants and manufacturers throughout the world use Group I's currencies for invoicing exports (see Sven Grassman); some are actively traded in Euromarkets as the basis for international borrowing and lending by nonresidents. By either standard the currencies of Western Europe, Japan, North America, and one or two others are potential reserve assets for other central banks.

*Department of economics, Stanford University. More support for the ideas in this paper can be found in my 1979 book.

The second group of more than a hundred nonindustrial economies is composed mainly of socialist countries and *LDC*s whose own currencies are either formally inconvertible or little used internationally—"minor" currencies in the terminology of Stephen Magee and R. K. S. Rao. Group II's foreign trade and exchange reserves are denominated in monies issued by Group I. Hence, the current world monetary order requires that the financially mature economies of Group I observe Article VIII of the International Monetary Fund: the commitment to convertibility. World trade would be undermined if the major industrial economies abrogated Article VIII, and with it the monetary services each provides to the rest of the world.

Among the N currencies from Group I in which world trade is invoiced, private commercial banks use the dollar as the common *vehicle* currency for clearing international payments in both spot and forward transactions. More than the efficiency of private money changing is involved, however. Central banks from Group I continue to intervene heavily in the foreign exchange market, although many have no formal parity obligations (see John Williamson). To avoid national conflicts, particularly under floating exchange rates, the dollar remains the common *intervention* currency.

In a world of N currencies, only $N-1$ governments can each have *one* independent intervention toward an exchange-rate target —either implicit or explicit. Consistency is most easily achieved if the $N-1$ official interventions take place against a common intervention currency, and the government of the Nth country remains relatively passive. Since World War II, the United States has stayed passive against dollar interventions by foreign governments. The Federal Reserve Bank of New York may support the foreign exchange policies of one or two major trading partners, for example, to help

the Bundesbank prevent the dollar from appreciating against the deutsche mark. Because of close American consultation with the Bundesbank, however, this does not constitute an *independent* intervention.

As a quid pro quo for this American passivity, foreign central banks automatically finance *U.S.* "deficits." Governments from Group I accumulate U.S. Treasury bills and bonds in direct proportion to their official purchases of dollars; those from Group II select a broader portfolio including Eurodollars and some other Eurocurrencies. Indeed, both groups have a growing demand for gross liquid reserves in these forms, often acquired by borrowing at long term in New York or in the Euromarkets. The IMF calculates that, exclusive of nonliquid gold and Special Drawing Rights (*SDR*), more than 80 percent of all official reserves were in dollar assets at the end of 1978.

What other nation—or group of countries—could replace the United States in its key currency role? To give other countries sufficient latitude to adjust their exchange rates and international payments without conflict among themselves, the *N*th country must be large in both its *GNP* and foreign trade. The capital market of the *N*th country should be broad and deep so foreign governments can comfortably hold exchange reserves in easily marketable nonmonetary debt (see Charles Kindleberger). Then drawing down (or building up) these reserves need not disturb the monetary system of the *N*th country itself, as is largely true of transacting in *U.S.* government bonds. A cornerstone of the international system is, therefore, that the American capital market remains open to foreigners.

II. The *SDR* Substitution Account and the Principle of Nondiversification

American monetary policy remained relatively stable from the 1950's to the mid-1960's: the *U.S.* Wholesale Price Index increased at less than one-half of 1 percent per year. If the anomalous position of gold had been resolved by official demonetization, with continued American monetary stability the strong dollar standard could have survived by nations *voluntarily* playing by the rules. For convenience, each foreign central bank would intervene only in dollars, hold dollar reserves because their value was stable, and voluntarily fix its exchange rate with the dollar to facilitate stabilizing the commodity purchasing power of its own currency.

However, in the early 1970's, the dollar standard changed from a strong to a weak form precisely because American monetary policy became unstable. After the advent of floating by major industrial economies to insulate their domestic monetary systems from American inflation, the convention of holding foreign exchange reserves in dollars—and these as direct claims on the United States—is endangered. First, continual erosion in real purchasing power of the dollar is inadequately offset by higher rates of interest on *U.S.* government securities. Secondly, great cyclical movements in exchange rates among convertible currencies—as in 1971, 1973–74, and 1977–78—provide strong incentives for foreign central banks to diversify. Indeed, although the dollar is still the most important currency, countries from Group II are now quite uninhibited in their selection of bonds or Eurodeposits in virtually any convertible currency. But Group I itself must *not* diversify: official exchange reserves of each industrial country should remain direct dollar claims on the United States. Why?

Keeping the intervention currency and reserve asset the same imposes an important discipline: only the exchange rate of the intervening country is directly affected. In contrast, suppose the Reserve Bank of Australia held Swiss francs and Japanese yen as reserve assets, with dollars remaining the intervention currency. To defend the foreign exchange value of the Australian dollar would necessitate first selling francs or yen for American dollars, thus upsetting the Japanese and Swiss exchange markets. Pure portfolio reshuffling by the Reserve Bank of Australia creates similar difficulties, violating the important principle that each central bank have only one independent intervention.

But why should official reserves be direct dollar claims on the United States rather than Eurodollars? If financial leakage from the unregulated Eurodollar system is high, the *U.S.* Federal Reserve can better control the total stock of dollar liquidity in the world. Similarly, foreign central banks from Group I can avoid undue pyramiding of their dollar reserves, leading to a loss of domestic monetary control (see Fritz Machlup), if officially acquired dollars are not redeposited in the Euromarket. Serious pyramiding in the late 1960's led the principal central banks from Group I to agree in June 1971 to stop acquiring Eurodollar deposits. Subsequent reserve acquisitions have been direct claims on the United States, and these captive foreign buyers have sometimes driven yields on U.S. Treasury bills significantly below those on Eurodollars.

To strengthen the 1971 agreement and to dampen current pressure on central banks to diversify out of *U.S.* government securities in the open market for foreign exchange, the *SDR* Substitution Account was recently proposed. The IMF would buy (and hold) *U.S.* government securities from foreign central banks in return for Account *SDR*—whose value is an index of the sixteen most important currencies used in foreign trade. In effect, holders of Account *SDR* would receive an *exchange-rate guarantee* against the dollar depreciating vis-à-vis other convertible currencies. Although details of Account *SDR* are not yet settled, my analysis suggests that

1) The Substitution Account be an official instrument among the IMF and participating governments without being traded on private markets.

2) The IMF buys U.S. Treasury bills and bonds, from governments other than the United States, in exchange for Account *SDR*.

3) Account *SDR* be traded freely among member governments, or sold back to the IMF for dollars less a small service charge. Net pressure for resale might arise from future surpluses in the American balance of payments.

4) Interest on Account *SDR* be paid in dollars at 100 percent of the average interest earned on the current *SDR* index of foreign currency government bonds. Presently, this interest is less than that earned on U.S. Treasury bills.

5) The *U.S.* government assumes full equity participation in the Account: i) in making good account losses if interest paid exceeds interest earned, while retaining any profits; and ii) in paying into the Account the equivalent of any net dollar depreciation against *SDR*.

6) Participating foreign governments agree not to further diversify their official exchange reserves into nondollar convertible currencies or gold in the open foreign exchange market.

Thus the IMF acts as a financial intermediary to guarantee exchange rates on U.S. Treasury obligations held by foreign governments. Because nominal interest on Account *SDR* is adjusted downwards from that on U.S. Treasury bills, which already anticipates some devaluation, the American government is given a fair bet. Although the idea of the *SDR* Substitution Account is to hold fast the main central banks from Group I to not openly diversifying, overall stability would be enhanced if central banks from Group II also participated. It is neither possible nor necessary to control private holdings of foreign exchange in the same way.

III. Internationalizing American Monetary Policy

Under the fixed exchange rates of the 1950's and 1960's, the money supplies of industrial economies other than the United States were largely determined by the balance of payments. Hence, in order to avoid indeterminacy in the world's money stock and price level, it was appropriate for the *U.S.* Federal Reserve to control *autonomously* the domestic supply of base money in dollars *without* reference to the *U.S.* balance of payments (see my 1969 paper). This policy basis of the strong dollar standard had two important facets:

i) The effect on the American monetary base of frequently building up or drawing down of dollar reserves by foreign central

banks was completely sterilized; only outstanding U.S. Treasury bonds and bills, but not depository claims on the Federal Reserve Bank itself, were affected.

ii) The *U.S.* government had a strong commitment, and sufficient financial control, to independently stabilize the dollar's purchasing power over internationally tradable goods and services.

True, automatic sterilization under (i) greatly accentuated the multiplier effect of any increase in the *U.S.* monetary base on the world's stock of money aggregated over all currencies (see Alexander Swoboda). But the inherited stability in price level expectations, coupled with a fixed exchange rate, served to stabilize the demand for base money (and for M_1) in dollars. Thus, under the strong dollar standard, American monetary authorities could tailor the supply of domestic base money to the demand for it.

Under the present weak dollar standard, what modifications in those Federal Reserve operating procedures are called for? How might the exchange-rate turbulence described by Jacob Frenkel and Michael Mussa be dampened, while domestic monetary stability is enhanced?

Unstable price inflation in the 1970's and fluctuations among convertible currencies have led to extrapolative, rather than regressive, expectations about further changes in exchange rates. These in turn can destabilize the demand for noninterest-bearing dollar balances—including the derived demand for base money. During the prolonged fall of the dollar through 1977 to November 1, 1978, a remarkable switching out of dollars into deutsche marks, yen, Swiss francs, and so on occurred. This switching reduced the demand for *U.S.* monetary aggregates substantially below the rate of growth in nominal American *GNP*. Thus monetarist rules, which presume stability in the demand for money, failed to predict the resulting inflationary explosion. The once-useful strategy of operating American monetary policy independently of foreign exchange considerations has become obsolete.

How then does one substitute benign attention for benign neglect? A credible commitment to stabilize the domestic and international purchasing power of the dollar is very important. The Federal Reserve's draconian measures of October 6, 1979 to slow growth in domestic bank credit is one correct step. Another would be to give an exchange-rate guarantee to foreign official holders of dollar balances by the United States taking a full equity position in the *SDR* Substitution Account.

In addition, the mechanics of short-run control over the American money supply should be reorganized. The recent abandonment of an official interest rate target on federal funds was long overdue. When dominated by expected changes in prices or exchange rates, nominal rates of interest are worse than useless targets or indicators for monetary policy; they led the Federal Reserve badly astray during the 1977–78 fall of the dollar (see my 1978 paper). While appropriate in the long run, by itself a simple Friedman rule of smooth growth in the American monetary base is similarly suspect. Instability in money demand is now rife because credibility in the rule is lacking, international currency substitution has become commonplace, and fringe banks and money-market funds are offering competing forms of money. Therefore, an additional governor on the short-run rate of base money creation in the United States would seem warranted.

Whether the dollar is "strong" or "weak" in the foreign exchanges is one immediate and potentially reliable source of information for the Federal Reserve to judge whether the American money market is unduly "tight" or "easy." On November 1, 1978, the American government finally recognized the importance of the foreign exchanges by intervening massively to halt the dollar's twenty-month slide *and* by allowing the American monetary base to contract commensurately with this intervention. The postwar policy of benign neglect was abruptly terminated. But how should the Federal Reserve process foreign exchange information, including the size of dollar transactions by foreign central banks, on a regular noncrisis basis?

Suppose 140 central banks from Groups I and II are actively buying or selling dollars.

Most countries have small dependent monetary systems, or have inconvertible currencies, or have greater financial instability than the United States itself. Hence the information contained in the official interventions of, say, 130 of the 140 countries is best disregarded by American authorities. Even coordinating *U.S.* monetary policy with as many as ten trading partners—albeit with convertible currencies and fairly stable monetary systems—is too complex: it departs too much from the *N*th country principle. Instead, consider gearing American monetary policy only towards Japan and Germany, the latter representing a European bloc. They are convenient benchmarks because both have large developed financial systems and presently display greater monetary stability than the United States.

Let the German and Japanese governments continue daily or weekly interventions to smooth the deutsche mark/dollar and yen/dollar rates to maintain a rough purchasing power parity. Then let the Federal Reserve Bank, the Bundesbank, and the Bank of Japan *mutually* adjust their domestic money supplies by the amount of these official interventions (see my 1974 paper). For example, if the dollar is sufficiently weak against the deutsche mark to warrant dollar purchases by the Bundesbank, the market is signaling that private demand for dollars has fallen and for deutsche marks has risen. Hence, the authorities should reduce the supply of base money in dollars and, symmetrically, increase the German money supply. In the short run, all three governments could better tailor their money supplies to shifting cross-currency demands, thus enhancing *domestic* financial stability in each country. Moreover, the possibility of a foreign exchange crisis would be drastically reduced.

To better manage this monetary accommodation on the American side, Germany and Japan could keep working dollar reserves directly in interest-bearing deposits with the Federal Reserve Bank of New York. (The other 138 countries would continue holding their main reserves only in nonmonetary U.S. Treasury bonds and bills with automatic sterilization of any impact on the American monetary base.) Hence, an intervention by, say, the Bank of Japan to buy yen and sell dollars would increase the American monetary base—as dollars are switched out of the Federal Reserve into American commercial banks—paralleling a contraction in the Japanese monetary base. Full sterilization—automatic or discretionary—of these foreign exchange transactions would contravene the agreement.

Friedman rules delimiting long-run *domestic* credit expansion by each of the three central banks are also essential for the success of the tripartite arrangement. The long-run rate of growth of nominal *GNP* in Japan would seem higher than that of the United States if both countries strove for stability in the yen/dollar prices of a broad basket of internationally tradable goods. Hence, domestic credit expansion by the Bank of Japan—through open-market operations or lending to commercial banks—would exhibit higher growth, say 9 percent per year against 5 percent for the United States, with Germany somewhere in between. Long-run monetary growth of the triumvirate would then be anchored, but short-run changes in national money supplies would be dominated by the foreign exchanges. Once the foreign exchange market perceived that the triumvirate's domestic money supplies were adjusting appropriately to shifting international demands, the need for official stabilizing intervention would diminish.

REFERENCES

J. A. Frenkel and M. L. Mussa, "The Efficiency of Foreign Exchange Markets and Measures of Turbulence," *Amer. Econ. Rev. Proc.*, May 1980, *70*, 374–81.

Sven Grassman, *Exchange Reserves and the Financial Structure of Foreign Trade*, Lexington 1973.

Charles Kindleberger, *The World in Depression: 1929 to 1939*, Englewood Cliffs 1974.

S. P. Magee and R. K. S. Rao, "The Dollar in International Trade," *Amer. Econ. Rev. Proc.*, May 1980, *70*, 368–73.

Ronald McKinnon, "Private and Official International Money: The Case for the Dollar," *Essays in International Finance*, Princeton Univ., No. 74, 1969.

______, "A New Tripartite Monetary Agreement or a Limping Dollar Standard?," *Essays in International Finance*, Princeton Univ., No. 106, 1974.

______, "Exchange Rate Instability, Trade Imbalances, and Monetary Policies in Japan, Europe and the United States," paper presented to Conference Honoring Harry Johnson, Oxford, England, Sept. 1978.

______, *Money in International Exchange: The Convertible Currency System*, New York 1979.

F. Machlup, "The Magicians and Their Rabbits," *Morgan Guaranty Survey*, New York 1971.

A. Swoboda, "Gold, Dollars, Euro-Dollars, and the World Money Stock under Fixed Exchange Rates," *Amer. Econ. Rev.*, Sept. 1978, *68*, 625–42.

J. Williamson, "Exchange-Rate Flexibility and Reserve Use," *Scand. J. Econ.*, No. 2, 1976, *78*, 327–39.

Structural Change and Regulatory Reform in the Utilities Industries

By HARRY M. TREBING*

There is growing disenchantment with the adequacy of the regulatory response to change in the electric, gas, and communications industries. Much of the current literature, however, fails to provide a viable program for regulatory reform. The capture theory of regulation would seem to indicate that the only option is to abandon regulation in favor of some type of auctioning or bidding process, or a system of excess profits taxes—alternatives which also suffer from significant imperfections. The coalition-building theory, which argues that regulatory institutions promote political coalitions and constituencies in order to survive and prosper, especially by programs of internal cross subsidization, offers little basis for optimism. It suggests that any efforts to reform regulation or strengthen the capability of commissions will simply result in greater cross subsidization in favor of a particular constituency. The equity-stability theory argues that legislatures promote regulation not because markets are inefficient, but because legislatures seek to protect society from the unimpeded operation of market forces. As a consequence, equity-fairness goals take precedence over efficiency-oriented solutions, but the formulation of the theory remains sufficiently vague to leave both the direction and the consequences of reform largely indeterminate.

The position taken in this paper is that regulatory reform can best be achieved by recognizing the infirmities inherent in the traditional application of price-earnings controls, especially when such controls are applied in a narrow context that ignores market structure considerations. By broadening the perspective of regulation to include structural and distributional considerations, a new set of pressures can be introduced which will enhance industry performance and augment the effectiveness of the regulatory process.

I. Traditional Model of Public Utility Regulation

Public utility regulation is premised on the belief that price-earnings controls can realize gains from economies of scale in production, while negating monopolistic distortions of price, output, and income. There are two assumptions implicit in this regulatory model. The first is that price-earnings controls can be applied with an acceptable degree of accuracy, rigor, and consistency. The second is that regulatory application of these controls will induce a pattern of behavior on the part of the regulated firm which is similar to the response of a competitive firm to market pressures.

A substantial amount of historic evidence casts doubt on the validity of the first assumption. Commissions have struggled to develop an elaborate system of rate base regulation (*RBR*) to determine revenue requirements. However, the system is still unable to resolve major disputes over the proper treatment of significant cost components, such as plant under construction or the rate of return on equity capital. Further, *RBR* does not respond well to significant changes in economic conditions. In periods of rapid market growth and potential excess profits, *RBR* creates an incentive to overinvest, maintain excess capacity, and introduce Averch-Johnson input distortions. Equally important, rate base expansion can

*Michigan State University.

camouflage earnings so that regulatory agencies cannot rely on the overall rate of return as a measure of success in controlling monopoly profits. These infirmities are further aggravated by an inability to establish standards for reserve margins or reliability levels or to integrate interfirm performance comparisons into the rate-making process. In periods of sharply rising costs, *RBR* introduces automatic adjustment clauses which permit the rapid pass through of cost increases to the consumer without formal rate proceedings. This creates another set of incentives for X-inefficiency and weakens regulatory surveillance. In addition, primary reliance on *RBR* in periods of price inflation leaves the regulator with few options other than to audit expenses and speculate over the magnitude and distribution of rate increases.

The situation is not significantly better with regard to rate structure reform. Despite the popularity of time-of-use rates, a debate still rages over what is an appropriate costing methodology. At present, at least four fundamentally different methods for measuring the marginal cost of electricity supply are being aggressively promoted by major consulting firms. The Federal Energy Regulatory Commission (FERC) has avoided this issue by the simple expedient of promulgating reporting requirements that only serve to build up a large data base which hopefully will be of value to someone. Pricing indeterminacy is further reinforced by the need to reconcile peak prices reflecting current costs with a revenue requirement reflecting an averaging of historic and current costs. Solutions such as the inverse elasticity rule appear to require too much information on cross elasticities over time, and are vulnerable to manipulation by the firm. Finally, pricing is complicated by the absence of generally accepted guidelines for implementing equity criteria, leading to a case-by-case approach.

A more serious shortcoming is inherent in the second assumption: that price-earnings regulation can influence the behavior of the firm just as competitive market prices discipline the competitive firm. The competitive market introduces a set of incentives and penalties as well as threats in the form of new entry and bankruptcy which are lacking under regulation. For the most part, commissions remain indifferent to matters pertaining to market structure, and no commission has ever followed a conscious policy of introducing bankruptcy as a means of promoting efficiency. As a consequence, the regulated firm has a significant degree of freedom to pursue its individual goals. Paradoxically, this freedom will be increased the greater the monopoly power of the firm.

II. Strategies of the Regulated Firm[1]

Assuming that the firm accepts some form of price and earnings control as a condition for public utility status, it will seek to assure 1) a high degree of corporate hegemony, 2) the highest long-run rate of return permissible under regulatory constraints, and 3) risk minimization. To achieve these objectives, it can employ a number of strategies.

To maintain corporate hegemony, the firm can seek to control market forces in a fashion that circumscribes rivals and forecloses entry. This can be accomplished by employing limit pricing, restricting access to the transmission and distribution networks, and establishing market-sharing practices.

There is ample evidence of the use of price to promote the hegemony of the firm. The Bell System's Telpak rates were set to foreclose private microwave systems, and Bell's rates for Dataphone Digital Service were set at levels which forced DATRAN into bankruptcy, even though these rates were later found by the Federal Communications Commission (FCC) to be unlawful. In electricity supply, "price-squeeze" cases provide evidence of the use of price by large, integrated utilities to constrain the ability of rival distribution systems to attract

[1]For a more detailed discussion of the strategy of the regulated firm, the benefits and costs of market structure reform, and a program for implementation, see my book with David Schwartz and Rodney Stevenson.

new industry to their service areas. Transmission rates for wheeling power also provide evidence of the use of price to segment markets and foreclose entry. Furthermore, opportunities for employing price to promote the hegemony of the firm will continue to exist in the future. The rates for access to exchange telephone plants, the rates for utility backup capacity for solar installation, and the rates for the output of cogeneration plants are cases where price can be employed to maintain the position of the established firm. At present, network access pricing is being resolved through bilateral negotiations among oligopolists, solar backup rates are largely unexplored, and rates for cogeneration are determined by the utility acting as a monopsonist.

A number of market structure practices have evolved which also serve to maintain the hegemony of the firm. In electricity supply, there are examples of territorial allocations between firms, refusals to deal—or, more correctly, refusals to sell at the wholesale level to small distribution systems, refusals to wheel power, and tying arrangements which limit the choice of utilities without generating capacity. Furthermore, power-pooling arrangements tend to cartelize markets, with restrictive conditions for membership which often exclude small firms. Conditions for pool membership may also require that participants accept territorial allocations, and agree to special terms favoring purchases and sales to other pool members, thereby limiting interpool transactions.

In the natural gas industry, the need to secure new gas supplies coupled with FERC indifference to market structure has served to promote vertical integration as both pipelines and distributors engage in joint ventures with producers to explore for gas. As a consequence, the arms-length dealings which prevailed between producers, pipelines, and distributors in the past have been replaced by a new community of interest. Joint ventures may also be expected to proliferate as the industry joins forces to finance high-cost synthetic fuel projects and the Alaskan gas pipeline. In common carrier communications, the FCC sought to promote competition in the terminal equipment and specialized carrier markets, while maintaining monopoly in the *MTT* and *WATS* markets. However, a new set of structural strategies has emerged as a consequence of the Execunet decision. The prospect of unlimited entry creates a strong incentive for communications utilities to establish unregulated subsidiaries that would be potentially free to enter all markets. In the case of the Bell System, such a subsidiary would begin business with an estimated sales volume in excess of $1 billion per year, and would give rise to major questions about cross-subsidization and market power.

The regulated firm may also seek to soften and dilute the effectiveness of the regulatory process to assure its continued dominance. This may be accomplished by lobbying activities, controlling access to information needed by regulators, employing delaying tactics in proceedings, and resisting policy changes that would affect structure and entry. Insofar as the latter is concerned, one can clearly discern AT&T's efforts to foreclose entry into telecommunications by setting forth successive arguments designed to show that entry was contrary to the public interest. Initially, the case rested on natural monopoly. This gave way to arguments based on systemic integrity, and, more recently, the question of the consumer's economic right to telephone service has been raised. Electric utilities, seeking to resist liberalized interconnection and pool participation, also make use of the systemic integrity argument.

To maintain long-run profitability, the firm again has a number of alternatives. It can exploit the vagaries of earnings controls while expanding the rate base to assure earnings growth—a practice that will be facilitated to the extent that pricing and market structure policies reinforce the monopoly status of the firm.

To minimize risk, the firm may promote those policies and practices which shift risk forward to the consumer. It can be argued that this is appropriate because regulation has limited profits over time to the benefit of the consumer, so the consumer should bear all risks associated with any major changes. The firm will be assisted in this process by *RBR*. Rapid cost increases will

be passed forward through automatic adjustment clauses, thereby minimizing the risk of regulatory lag. The burden of excess capacity, redundant plant, and gas curtailments will be shifted forward as *RBR* maintains the level of operating income in the face of diminished sales. Similarly, *RBR* will shift the cost of pollution abatement forward regardless of the success or failure of the technology employed; even effluent charges will be shifted forward, thereby negating any incentive to select least-cost solutions to pollution problems.

The regulated firm may also seek to employ price to shift risk forward to the consumer. When incremental costs are rising sharply, as is the case for new high-cost supplemental gas supplies, the firm will advocate rolled-in or average cost pricing. When incremental costs are falling, as is the case for segments of the telecommunications industry, the firm will advocate incremental pricing. The former tends to minimize the risk of selling the output; the latter tends to foreclose competitive entry. The FERC has adopted rolled-in pricing for liquified natural gas. Congress has adopted a highly limited form of incremental pricing for high-cost gas sales to boiler fuel users, but once the incremental price reaches the price of alternative fuels, rolled-in pricing is reinstated. In the Great Plains Coal Gasification Project, the FERC adopted an "all events" tariff which provides the ultimate assurance that risk will be shifted forward. The customers of the pipelines participating in this project will reimburse virtually all costs of the project even if it should fail.

In such a regulatory setting, the principal countervailing force against the employment of these strategies will be a commission's effectiveness in applying price-earnings controls. Even if the commission's resources are increased, the results will be no better than the adequacy of the controls, and as we have seen, they are subject to severe shortcomings.

III. An Alternative for Reform

A public policy option for reform could involve broadening the perspective of commissions to recognize the contribution of market structure toward improved industry performance. This would require reorienting regulatory intervention to focus on market structures that are conducive to collusion or subject to cartelization. It would require commissions to ascertain the structural ramifications of pricing practices and the neutralization of monopoly focal points (such as the transmission function where scale economies limit the number of suppliers) to assure free access. It would also require commissions to determine where competitive markets might be established even though they do not exist at the current time because of blockaded entry, high transaction costs, or high information costs. In such cases, the agency could recommend that new institutional arrangements be created to facilitate market transactions by performing brokerage functions and arranging for coordinated operations. To the extent that such markets can be created and restrictive practices curtailed, a new set of pressures for greater efficiency will be set in motion and the discretionary market power of the firm will be significantly limited.

The wholesale market in the gas industry and the bulk power market in electricity supply are excellent candidates for introducing this type of market structure policy. An interconnected transmission network would permit buyers and sellers to trade in gas rights and pay for varying degrees of firm or interruptible service through contractual arrangements. This would tend to alleviate demand-supply imbalances. More important, free access to a regional or national market would provide a strong stimulus for new entry into the supplemental gas market. Those entrepreneurs who are successful in gasifying coal, locating new gas supplies, or negotiating for the importation of *LNG* would now find a greatly expanded market available to them for their output—particularly if they were able to deal with both distribution systems and ultimate consumers. Initial efforts were made to create a so-called "white market" for the transfer of gas among different customers in New York State, but this proved unsuccessful because the size of the market as well as the means to effectuate transfers were severely limited. To achieve a competitive wholesale gas

market, it would be necessary to have an interconnected pipeline-distribution network with common carrier status, a series of transmission prices which represent distance and capacity utilization, and a market mechanism to perform the brokerage function.

Bulk power supply provides an even better example of the potential use of this type of market structure policy. Access to bulk power is limited by access to the transmission network, and entry is possible only by participation in a power pool or by arriving at a bilateral agreement with another utility system to interconnect. Reconstituting bulk power supply as a competitive market with free access would greatly expand the number of buyers and sellers. Suppliers could include existing utilities, new generating capacity built on a joint venture basis, and new entrants such as firms engaged in cogeneration. By expanding the scope of the bulk power market, firms would be better able to achieve high capacity factors because surplus power could be sold expeditiously through the market. An increase in the total number of suppliers would also permit more efficient operation of existing generating plants in the short run by employing short-run marginal cost pricing. In the long run, the market would permit cost minimization by offering each firm the option of either contracting to buy power or building additional capacity. An expanded market would also lead to greater exploitation of the economies of scale inherent in high-voltage transmission and minimize the opportunity for price discrimination between different classes of consumers. A bulk power market would also be of considerable value to state commissions in judging whether the least-cost method of expanding capacity will be utilized.

It is doubtful whether a policy of gradualism will bring about a competitive bulk power market along the lines discussed. For most electric utilities the transaction costs and information costs would be so great as to preclude initiating such a comprehensive interconnected network. Transmission requirements would have to be developed sequentially between firms on a bilateral basis to permit power exchanges outside of the firm's service area or outside of an established pooling arrangement. Furthermore, the regulated utilities would be reluctant to develop such an arrangement for fear of losing a portion of their rate base or their wholesale power markets.

There could be substantial external benefits associated with the creation of new, expanded wholesale gas and bulk power markets. For example, these markets could increase the number of power plant sites available, and by providing new sources of gas permit the construction of fossil fuel plants in nonattainment areas, thereby opening up still more sites without altering pollution standards under the Clean Air Act. An expansion of the bulk power market could also serve to improve reliability at minimum cost because of the diversity of the expanded range of generation sources. In addition, both markets would tend to stimulate and decentralize research and innovative activity.

A regulatory initiative along these lines seems warranted, particularly as shortages grow in the energy field and the distinction between monopoly and competition becomes blurred in telecommunications. There will still be important problems that these reforms do not address directly, such as the distributional consequences of rising energy prices and the existence of uncompensated social costs. However, the control of restrictive practices and the creation of competitive markets may tend to narrow the range of these problems.

REFERENCES

David Schwartz, Harry Trebing, and Rodney Stevenson, *Competition and Regulatory Reform in the Energy Utilities*, forthcoming.

Radio Spectrum Allocation: Role of the Market

By WILLIAM H. MELODY*

The radio spectrum has historically been viewed as a scarce natural resource to be allocated by national governments and international agencies rather than by markets. As demand for increased use of the spectrum has grown dramatically over time, the economic value of major portions of the spectrum has also increased dramatically. In turn, the problems and costs of congestion and interference have increased significantly. Evidence has been uncovered demonstrating inefficient use of portions of the spectrum as measured in traditional economic terms. An administrative process that allocates valuable spectrum without charging a "price" to users has come to be recognized as one that provides incentives to promote the wasteful use of the spectrum resource and to encourage uneconomic stock-piling of spectrum licenses.

For more than a decade, professional journal articles, studies and reports, in the United States, Canada, and other Western countries have addressed various aspects of the problems due to not recognizing economic factors in the process of allocating the radio spectrum. The issue was discussed in the 1968 President's Task Force on Communications Policy. It was discussed in Canada in *Instant World*, 1971. More recently, it was discussed in the 1977 *Option Papers* of the U. S. House Subcommittee on Communications as part of its reconsideration of the 1934 Communications Act. Presently, both the Federal Communications Commission in the United States and the federal Department of Communication in Canada are considering making greater use of fees and charges in their spectrum allocation policies.[1]

Suggestions from economists for modifying the existing administrative process of spectrum allocation in North America range from the incorporation of market criteria into the administrative allocation process to the substitution of private market allocations for the administrative process. At the same time, interest in the radio spectrum has grown to include developing nations which now represent a majority within the International Telecommunications Union. The World Administrative Radio Conference met for the first time in twenty years in fall 1979 to consider revisions of past policies and practices followed in administering the spectrum. Technical sessions for engineers were expanded to encompass not only economic issues, but political and social issues as well.

I. Radio Spectrum Characteristics

The radio spectrum is a unique natural resource that differs from other resources in certain fundamental ways. Interdependence among users is extremely high. Cooperation among all users is essential if the spectrum is to be used effectively by anyone. Possibilities for negative externalities through interference are very great. Nationwide and worldwide cooperation exist now, and has been maintained by opposing sides even during wartime. Although interference can reduce effectiveness and efficiency, spectrum capacity cannot be exhausted. In fact, with technological advance, spectrum capacity has been enhanced continuously.

Measurement of use of the spectrum, or even of rights to use it, cannot be objectively specified. Rather, spectrum rights are probabilistic in nature depending upon environmental conditions at the time and the use of the spectrum by others. Rights to use the spectrum are not susceptible to legal enforcement as are private property rights. In the past, allocation by the market of

*Department of communication, Simon Fraser University.

[1]See also Canada, Department of Communications, 1976; Harvey Levin; J. O. Robinson.

rights to use the spectrum has been found to be impossible, or inefficient. The spectrum has been recognized as a social resource, by both domestic and international law, a unique form of social property.

The history of the radio spectrum has been tied directly to military interests, in terms of *R&D* to expand the capacity of the spectrum and dominant user assignments. More recently, because the spectrum is the building block of most other communications technologies, its crucial significance affecting the flow of information, which forms the basis of political power, has been recognized by Third-World countries. Internationally, spectrum allocations have acquired enormous political as well as economic significance.

In North America it is not difficult to find examples of apparent inefficiency in the existing system of spectrum allocation. There is congestion in some bands and idle space in others. In some instances, spectrum assignments are "banked" for future use simply because there is no cost involved and competition for the assignment would likely be more severe at a later date. Decisions are often made to use the spectrum instead of substitute technologies (for example, cable) because the cost of the spectrum is near zero. The design and use of equipment is frequently based on the use of relatively large portions of the spectrum so as to conserve on material and labor inputs to the equipment manufacturing process.

Finally, under the existing system of administrative allocations, where indirect market exchanges do take place, the substantial monopoly value of the social resource is appropriated by private interests which are in a position to exploit the major imperfections in the market place created by the characteristics of the spectrum and the administrative process. For example, commercial broadcast licenses are assigned by administrative decision. Recipients of broadcast licenses generally receive a substantial economic value because of their monopoly right to that spectrum assignment. When the broadcast station and its license are sold in private markets to a new owner, the market price reflects the value of the spectrum resource in a marketplace where direct competition is severely restricted.[2]

II. Market-Based Proposals

The range of proposals to improve the efficiency of spectrum allocations includes: 1) creating a market in freely transferable spectrum rights to substitute for the administrative allocation process; 2) auctioning spectrum licenses; 3) establishing shadow prices for spectrum assignments and/or exchanges; 4) establishing administered license fees based on opportunity cost or other standards of valuation.

The suggestion to establish a market in spectrum rights was first made by Leo Herzel, and has been considerably developed by Ronald Coase, J. R. Minasian, and William Meckling. The basic rationale is found in neoclassical theory of competitive markets, as refined principally by Coase's analysis of the role and allocative function of property rights. The other proposals build on the same theoretical foundation in attempting to incorporate market criteria for spectrum allocation, with particular reference to market valuations of licenses.

Drawing from neoclassical market theory, this literature assumes that: 1) market-determined decisions are automatically superior to administered decisions; 2) any proposal to make greater use of markets is automatically a move toward optimal allocational efficiency, whether or not those markets would be workably competitive; 3) opportunity costs, however defined and measured, given any market structure, will automatically improve allocational efficiency; and 4) optimal efficiency as defined in neoclassical theory is, or should be, the objective of the spectrum allocation process. Unfortunately, none of these assumptions are true.

In addition, the fundamental requirement of any market in spectrum rights is that property rights must be defined and pri-

[2]In both Canada and the United States the transfer of licenses is subject to approval by the appropriate regulatory authority, the CRTC or the FCC. However, the conditions of the transfer do not include regulatory intervention in determining the transfer prices.

vately transferable. The probabilistic nature of emission patterns and the interference problems created as well as other inherent market rigidities raise serious doubts as to the possibility of satisfactorily defining private property rights in spectrum. For it is likely that transaction and enforcement costs as well as loss of spectrum use would far exceed any claimed allocative efficiency benefits. An outline of the operational characteristics of such a market has yet to be developed.

Even if one accepts the possibility of resolving the problem of spectrum rights definition, the desirability of a market system remains extremely questionable on other grounds: 1) Such a market system would fail to take into account the very substantial externalities associated with the provision of spectrum-using services. This divergence between social and private valuations of spectrum worth in a particular use implies that market allocation would be socially inefficient even when considered on its own terms. 2) The noncompetitive nature of the markets in which spectrum users operate implies a further divergence between social and private valuations of spectrum worth in particular uses. The implications in this regard are particularly serious when monopoly users regulated on a cost-plus basis are involved (for example, telephone common carriers, pipeline, railroad and electric utility firms). 3) Major users of the spectrum are government and other public agencies at local, state, and national levels, that neither dispense services nor attract capital through private markets. The nature of the fiscal budgetary systems for such agencies precludes them from equal market participation. 4) Administrative discretion would be severely narrowed. It would restrict, and could render the system incapable of achieving broader economic, social, and political objectives.

The above obstacles are sufficiently serious to preclude consideration of a market system in spectrum rights, without substantial further research, even if one views the objective of the spectrum allocation process to be only narrowly economic. The incorporation of economic criteria into the process of spectrum administration is worthy of more detailed investigation than has occurred to date. But the contribution will be limited as economic criteria facilitate rather than replace spectrum administration. The objective of the spectrum administration process is not simply to imitate a market, or to adopt the economic valuations that a market might yield. Given the essential characteristics of the spectrum, the perfectly competitive market model seems particularly inappropriate as a paradigm for defining an efficient allocation of the spectrum resource.[3]

III. Toward Improved Spectrum Analysis

A more relevant analysis of the spectrum is likely to be developed by building on other branches of theory. One is Ricardian rent theory and the vast literature addressed to the taxation of economic rent as the "unearned increment" from private ownership of land. There are many direct parallels in the analysis between these two natural resources.

The spectrum, like land, is not homogeneous in its productivity. There is a level of economic rent determined at the margin of cultivation. There is an intensive margin measured by the cost of expanding the communication capacity of existing radio frequencies. There is an extensive margin measured by the cost of making higher frequencies usable. The cultivation and expansion of spectrum productivity is governed by spectrum *R&D*. The latent communications capacity of the spectrum is vast, if society is willing to incur the costs necessary to make it productive. Yet, at any given time there is a general scarcity that is unevenly distributed throughout the spectrum. The uneven incidence of interference and congestion within the radio spectrum suggests the applicability of Ricardo's extensive and intensive margins of cultivation. Historically, the extensive margin of the spectrum has been pushed from low frequency to higher and higher frequency

[3]See my paper with Dallas Smythe.

bands. At the intensive margin, where congestion and interference have become intolerable, *R&D* has been directed to reduce it.

There is substantial economic rent being realized by many users of the spectrum. The beneficiaries of the economic rent have been determined by the administered frequency assignment decisions. Not all users are able to convert this economic rent directly into profit by selling their frequency licenses, but all users do benefit from the opportunity to employ the spectrum resource in their respective production processes at costs that are less than its economic value, including rent.

It is presently the practice in North America for the administrative authorities to charge fees for radio licenses, but these are nominal and generally limited to the costs of administration. This analysis indicates that authorities should explore the possibilities of developing fee structures that will capture for the public treasury at least some part of the unearned increment of economic rent.[4]

For those users who employ the spectrum but cannot sell their licenses directly, the problem of spectrum valuation is a difficult one. Several of the problems raised above in respect to neoclassical market theory will apply to rent theory as well. However, where there exists a market in licenses, spectrum valuations can be obtained readily. Thus, there is reason to focus attention initially on those license assignments where a market already exists, not those where markets do not exist. The prime candidate is the broadcast services which generate their substantial unearned increments almost exclusively from using the spectrum.

A second area of theory and practice directly relevant to spectrum analysis is that concerned with the efficient allocation and management of common property resources, including forests, fisheries, and other natural resources. Theory that directly considers interdependence, externalities, the need for sharing and compensation rules, and the necessity of total system management for system efficiency must represent more fertile ground for analysis than neoclassical market theory with its assumptions of independence and atomistic private markets.

The spectrum resource is a public commons. The right to enter and use the commons is presently governed by administrative authority. Under the existing institutional arrangements, there are some distorted economic incentives and inefficiencies. Improved efficiency requires that these distortions be examined within the context of the characteristics of the spectrum commons and the institutional relations surrounding its allocation and use.

IV. Conclusion

The spectrum will continue to be allocated and assigned by means of an administrative process. This is required by the characteristics of the resource. The market cannot be an efficient substitute for the administrative process in achieving either allocational efficiency or the broader objectives of the process. As the spectrum becomes more valuable over time, the costs to society of inefficiency in the process will increase, but the broader political and social criteria are likely to assume new importance.

The primary need is to recognize the administrative process as one requiring active planning and management, rather than technical management of licenses allocated on a first come, first served basis. Within such a program there is ample room for economic analysis to contribute to the establishment of standards and incentive systems as well as sharing and compensation rules. But economic contributions are not likely to come from attempts to turn the spectrum over to private economic markets under the belief that allocational efficiency somehow will be improved. They are much more likely to come from analysis building on the theory of economic rent and theories of common resource management. This is particularly so when it comes to developing the relations between economic and noneconomic objectives of the process. Theories

[4]See my paper with Smythe and A. Oliver.

of economic rent and common resource management can provide a basis for developing a system of compensation payments and sharing rules designed to achieve efficiency in light of the broader objectives.

Additional problem areas where economic analysis can usefully assist the spectrum administration process include: 1) specifying the relevant economic criteria for both the planning and implementation phases of the process, including evaluations of the economic implications of alternative criteria and the development of incentive systems and cost-sharing rules; 2) assessing the relations between *R&D* and the obsolescence created by it, including the structure of *R&D* and obsolescence problems at both the extensive and intensive margins of spectrum cultivation; 3) devising systems to facilitate adjustments to changing conditions, whether by administered exchanges, fees, markets, or other methods; 4) developing standards for bringing closer together the consequences of private economic behavior and system efficiency of the common spectrum resource.

For the radio spectrum, the market is one of many institutional tools for policy analysis. It cannot, and almost certainly will not be the policy objective.

REFERENCES

R. H. Coase, "The Federal Communications Commission," *J. Law Econ.*, Oct. 1959, *2*, 1–40.

______, **W. Meckling, and J. R. Minasian,** "Problems of Radio Frequency Allocation," unpublished study, Rand Corp., May 1963.

L. Herzel, "Public Interest and the Market in Colour Television Regulation," *Univ. Chicago Law Rev.*, Oct. 1951, *9*, 802–16.

Harvey J. Levin, *The Invisible Resource: Use and Regulation of the Radio Spectrum,* Baltimore 1971.

W. Meckling, "History of International Allocation and Assignment," paper presented at the Conference on Economics of Regulated Public Utilities, Chicago, June 20–25, 1965.

W. H. Melody, and D. W. Smythe, "A Study of the Feasibility of Applying the Opportunity Cost Concept to the Spectrum Allocation Process," rept. to Department of Communications, contract no. 0SU77-00368, Ottawa, Mar. 1978.

______, ______ **and, A. Oliver,** "A Study of the Applicability of Economic Cost and Value Criteria in Establishing License Fee Schedules for Radio Spectrum Assignments, with Particular Reference to the Microwave Range .890 MHZ to 16 GHZ," rept. to Department of Communications, contract no. 02SU-36100, Ottawa 1979.

J. R. Minasian, "Property Rights in Radiation: An Alternative Approach to Radio Frequency Allocation," *J. Law Econ.*, Apr. 1975, *18*, 221–72.

J. O. Robinson, "An Investigation of Economic Factors in FCC Spectrum Management," rept. no. SAS 76-01, FCC Spectrum Allocation Staff, Washington 1976.

Canada, Department of Communications, *Instant World*, Ottawa 1971.

______, *A Proposal for Restricting the Radio License Fee Schedule*, Ottawa, Dec. 30, 1976.

Federal Communications Commission, "Fee Refunds and Future FCC Fees," *Federal Register*, No. 196, Oct. 10, 1978, *43*, 46658–92.

U.S. Congress, House, Interstate and Foreign Commerce Committee, Subcommittee on Communications, *Option Papers*, Washington 1977.

U.S. President's Task Force, "The Use and Management of the Electromagnetic Spectrum," Parts I and II, staff paper no. 7, U.S. Department of Commerce, Washington 1969.

State Initiatives in State-Federal Relations

By THOMAS K. STANDISH*

Adjustments forced upon energy utilities by OPEC, challenges by IBM, ITT, and others to the traditional franchise rights of AT&T, the rapid rise of CATV toward public utility status, and deregulation moves in transportation signal an era of rising importance for the regulation of public utilities. Regulatory response to the demands of advocate parties will shape and form our economy in the 1980's and 1990's. Existing regulatory tools, methods, and concepts are not adequate to handle the challenges presented by these changes in the economy. Adapting to new operating conditions will require the use of dramatically new approaches to regulation. It is critical to know the form these solutions should take to best serve our future. Based upon historical analysis, this paper examines the proper role of regulation at the state, regional, and federal levels of government.

Four areas of inquiry are highlighted by the so-called "crises" in public utility industries. The first question is whether there should be a shift from passive to active regulation. Is there an increased need for government regulators to function as an investment risk diffuser, and is it prudent for our regulators to actively mediate conflicts between competing interests involved in the public utilities area? Or does the existence of the regulatory "crucible" encourage advocates to become more extreme. Second, is increased regulatory power needed at regional and federal levels of government in order to achieve an appropriate balance between state, regional, and federal government, or does overcentralization breed economic inefficiency and autocratic decisions? Third, should regulators have a higher degree of power and control over private utility companies to accomplish objectives mandated by state legislatures and Congress, or is the real problem that private incentives have not been given enough leeway to solve the problems which confront us? And, fourth, if we are to address today's opportunities, should planning efficiency and public interest norms displace the time-honored competitive market standards of economic efficiency and social justice as the basis for regulatory decisions? Each of the four areas of inquiry enumerated above, while topical today, are as old as regulation.

At the root of these issues, there is a clash of perspective, a clash between those arguing for the use of a planning paradigm to solve economic problems and those advocating reliance upon market forces. Curiously, the significance of this clash is only apparent. Economic imperatives lead us to expect a continuation of a trend which resolves the market/nonmarket balance in favor of more government participation. Hence, the scope of state initiatives and the range of opportunities for policy action will in fact be defined by economic forces larger than the current debate on the issues.

I. Efficiency Increase and Economic Scale

The drive for profit through increased efficiency has been at the core of the competitive market process. Innovations to introduce new products and to bring about lower costs have accounted for the increasingly high standard of living which underpins and stabilizes the social and political fabric of market economies. Historically, the forces which have brought about increased economic efficiency have been embraced as beneficial and, consequently, there has been little opposition to innovation. There have been, however, effects of rising efficiency which have significant implications for the organization of society. It is these feedback effects from increasing efficiency which have defined the role of the government sector in the economy and society. These

*Hartford, Connecticut

effects will then define the scope of state initiatives in federal-state relations.

Principal among the historical manifestations of innovation has been the increase in scale of operation of firms producing goods and services (see John Kenneth Galbraith, pp. 11–20). Scale increases in the firm have been paralleled by countervailing effects in the market for labor and, as argued here, by a government sector of ever-increasing size and regulatory power necessary to guarantee market stability. I turn now to an analysis of some aspects of increasing scale and the implications for initiatives in regulation.

Scale increases are of two types, those which affect the technical scale of production and those which affect the organizational-financial scale of the firm. Over the course of development of the market system, the effects of each of these types of change in scale has become the cause of future innovations. Of critical importance to the hypothesis offered here is to note that ever-increasing scale of economic production has also meant, by definition, a progress toward centralization of control, interdependence, and total system integration.

1) Technical scale has increased principally in areas where either labor is displaced by automation or new materials are introduced into production. That is, innovations affecting scale manifest an increase in the ratio of machine task to labor task, increased materials specificity, a wider materials diversity in production, and the like. These changes have been integrated with changes in the market for labor where human capital, reflecting increasing labor specificity and diversity, has acted as a complement to machine capital. Complex production has been facilitated by increased precision in machine and materials tolerances, more precise timing, and by refinements in the management decision process. The flow of production within industry has become progressively fused and centralized. Small-batch production has given way to large-batch production and then to process production (see R. Averitt, pp. 29–35), and where process production is not possible, the computer has permitted management to treat production as if it were a totally automated process unit. Hence, both the time horizon and the size of the management decision unit have historically increased—counter to Edward Chamberlin's expectation that management limitations would thwart economies of scale to the firm (pp. 235–36). Increases in information accuracy, speed of communication and, information complexity have reinforced this drive toward larger decision units and expanded management time horizons. Scale changes to increase efficiency have produced the centralizing of energy production and distribution.

Technological complementarity between and within production processes described above has meant that both the feasibility and the necessity for planning have arisen together. The need for long-run planning appeared first within the firms with large-scale production. Ironically, then, the source of expanding efficiency has had unintended results which have necessitated the gradual adoption of complete system thinking. This growth in feedback complexity has been exponential, not arithmetic, causing the accepted view of economic efficiency to expand from concern with efficiency within the firm to include a focus on total system efficiency. The primary emphasis has shifted away from private production to public imperatives which transcend market-determined objectives. This threshold was crossed within the last decade.

2) A rise in the organizational-financial scale of production has its origin in three principal sources—the need to finance ever-larger commitments of capital for technical scale increases, the drive to spread risk, and attempts to take advantage of monopoly power derived from financial centralization and/or market dominance. As a result of increased organizational-financial scale, a synergism has emerged wherein, in the place of smallness competitive with bigness characteristic of earlier stages in history, smallness is now both supportive and supported by bigness in market organization (see Averitt and my article). Increasingly, in prior production stages, in marketing dealerships and franchises, and in other forward stages of

vertical integration, small firms link with larger firms to knit an integrated fabric of industrial organization. Increased centralization has not been limited to the producer sector of the economy, rather, the financial sector has come to be dominated by larger organizational units which are interpenetrated with the producer sector. The trend is toward conglomerates which embrace all major sectors of the economy.

What emerges from an inspection of history is a picture of a highly complex and efficient economic machine with technical and organizational parts fused together into an integrated whole. Whole sectors of the economy increasingly function as one producer unit. Importantly, over time, component parts of the economic machine become more efficient through increased specialization and through interlocking with other elements in the system. This efficiency, however, has a price. Increasing specialization and interdependence between functional parts has produced rigidity and inflexibility in the economy. The very specialization and large scale necessary to achieve high productivity makes the market system vulnerable to breakdowns in key component parts of the system (see L. von Bertalanffy, essays 2 and 3).

II. To Counter Market System Rigidities, Government Must Act to Stabilize Operating Conditions and to Diffuse Risk

In the absence of stabilizing forces outside the market, the historical expansion of technical and organizational-financial scale would escalate the risk cost of entreprenurial decisions to prohibitive levels. This "eggs in one basket" phenomenon within the firm is magnified by the fact that the entrepreneur must have a degree of planning certainty, not just about conditions affecting his or her own production and markets, but also about the interlocking decisions of other firms and industries. Given that a few large-scale firms dominate most markets, a production or marketing decision by a firm in a competitor industry or by a major supplier has the potential to severely destabilize operations for the firm and, for that matter, for whole industries.

In sum and substance, efficient large scale introduces the potential for disruption both within and between industries. What is required to cope with these inflexibilities is the same as what is needed for planning—stable expectations over a long time horizon.

The planning function within the firm has shifted away from short-run concerns the producing and selling of output. Unlike his historical antecedent, today's entrepreneur focuses upon the long range and regards elaborate planning as an indispensable and intrinsic part of the decision-making process. The sheer scale of operation necessitates the adoption of a long-range planning mode within the firm (see Robert Heilbronner, 1970, p. 24), and in fact, scale increases within the firm expand and interlock with feeder industries. The severe risk to the economy presented by certain sectors is now being felt—it is becoming clear, however, that the obvious example of energy is but one instance of a more generalized phenomenon.

To cope with the resulting fused economic system and the exponential impact on investment risk from associated rigidities, business has sought government participation to stabilize operating conditions. In 1975, for example, Henry Ford II "called for the creation of a highly visible and vocal federal planning body, not because some wild radicals demand it but because businessmen will demand it to keep the system from sputtering to a halt" (see T. H. Naylor, p. 18). Business has sought government participation not just to stabilize their own investment environment as narrowly defined, but also to stabilize the operation of the interconnected system itself. In this sense, the Keynesian revolution came to be welcomed rather than resisted once the business sector recognized the role government could play to produce system stability.

The public sector, principally the regulatory function, is necessarily and appropriately a mediating influence which provides a stable context for economic development in a private sector facing ever-increasing investment risk. The gradual elevation of the importance of the government sector in the economy has in the past and will in the future be necessary to cut the growing

risks associated with expanding scale. (See Adolph Lowe in Heilbronner, 1969, pp. 167–99).

III. Active Interventionalist Regulation is Needed to Stabilize Public Utilities Sector

Utility regulation by the public sector has the function of stabilizing the operating environment for regulated industries. Because these industries are principal to the economy, stability and investment risk diffusion in the overall system results from regulation. Far from being a "natural" monopoly to be isolated in economic theory, the experience of public utilities points to the future for other industries which are moving toward advanced scale of production.

What this portends for the role of public utility regulation is a shift from a passive rate-of-return approach toward an active partnership with the private sector. This phenomenon has and will continue to appear first in the form of joint goal setting between the private and public sectors. Norms derived from concepts of planning efficiency, based upon pragmatic analysis, will replace the antiquated norms set forth in the traditional theory of price. Confronted with demands from an expanding number of advocates, public utility regulators will be forced to construct pragmatic efficiency measures unaffected with concerns such as Pareto optimality. Rather the aim of regulators will be to make the economic system work (see my article). Emphasis will continue to be placed on measures which stabilize the operating environment and thereby reduce private investment risk for utility companies. In this regard, public utility regulators will increasingly work in concert with other regulatory functions such as environmental protection, economic development, and energy planning to ensure that utility regulatory decisions dovetail with the requirements of the rest of the economy (for example, PURPA requirements; Connecticut Public Utilities Act of 1975). In addition, just as organizational-financial scale increases have led to significant efficiency gains in the unregulated sector of the economy, regulators will be required to permit utility mergers (including conglomerate mergers) as increases in risk continue to escalate for private utility companies. This will occur, however, only as fast as legislators and the economics profession abandon the notion that utilities are natural monopolies to be regulatorily isolated from the rest of the economy of which they are so obviously an integral part. Finally, the implications flowing from the analysis presented here point to a centralization of decision making at higher levels of government in order to be responsive to the needs of total system integration. Hence, we can expect not only more regulatory powers and control at all levels of government, but also a relative thrust toward expanded regional and federal regulatory powers. The tensions reflecting this thrust were reflected in the battle over the PURPA legislation and are exemplary of resistance to inevitable federal control over certain areas of energy rate making. In this vein, the moves toward deregulation currently being advocated are but a temporary pause in the drive toward increased regulation. In reality, as time passes it will be seen that moves to deregulate are aimed at the plethora of regulatory abuse rather than at the need for deregulation per se.

In summary, as a consequence of long-run forces which cause increasing scale of operation, a public-private partnership is emerging in the regulated sector. The apparent clash between the planning paradigm or the competitive market paradigm is yielding to the real world needs to make the economic system work. The result is that debate over the "issues" is being driven by an economic imperative away from traditional theory and toward pragmatic efficiency guides necessary to manage tomorrow's large-scale, interlocking economy.

REFERENCES

R. Averitt, *The Dual Economy*, New York 1968.

E. H. Chamberlin, "Proportionality, Divisibility, and Economies of Scale," *Quart. J. Econ.*, Feb. 1948, *62*, 235–36.

John Kenneth Galbraith, *The New Industrial State*, Boston 1967.

Robert Heilbronner, *Economic Means and*

Social Ends, Englewood Cliffs 1969.

———, *Between Capitalism and Socialism*, New York 1970.

T. H. Naylor, "The U.S. Needs Strategic Planning," *Business Week*, Dec. 1979.

T. K. Standish, "Integrating Efficiency Standards Into the Regulatory Process," in Harry Trebbing, ed., *Current Issues In Public Utility Regulation*, East Lansing 1980.

L. von Bertalanffy, *General System Theory*, New York 1968.

DISCUSSION

DALLAS W. SMYTHE, Simon Fraser University: The Trebing and Standish papers concern three questions: Is the process of regulating public utilities symmetrical as is usually assumed? Is it a viable institution under conditions prevailing in the 1980's? "Will we be on time for *1984*?"

Harry Trebing reviews the formidable problems facing the regulatory process, especially those stemming from its attempt to regulate price and earnings. It seems the problems in the traditional regulatory process put the regulatory authority in the position of the person on the treadmill. No matter how doggedly the runner persists, the treadmill looms ahead. The net effect is to put the viability of price-earnings regulation in question.

The viability question is sharpened if we reexamine the basic assumptions underlying it. At the outset, it was assumed that the regulatory institutions and courts would have power at least equal to and perhaps greater than that of the regulated firms. In reality this was never the situation. In terms of expertise (paid for as operating expenses passed on to the consumer), the regulated firms always outweighed the regulators. There has been a fatal ambivalence in the posture of the regulators because of the terms of the statutes and how they were construed. The regulator was supposed to simultaneously favor consumer interests while also being impartial in a quasi-judicial process. The statutes typically failed to confer power to regulate certain functions of the regulated entity which resulted in the latter presenting the regulators with *faits accomplis*. I refer here to innovation, financing, planning, and public relations (including advertising). Finally, there was a basic imbalance in effective political power. The monopoly to be regulated rolls on apparently permanently as a political power formation. But consumer advocacy is a transient, and unstable political force.

The whole apparatus was created because the private enterprises concerned wanted to head off public ownership between the 1880's and 1920's, and because naive reformers supported their initiatives. As a secondary advantage the regulated companies could use the apparatus to rationalize their structures and policies. Rate-base regulation, now nearing its first centenary, embodies an asymmetrical contradiction favoring the regulated companies which is why it seemingly is a permanent feature of the private power structure of America.

Thomas Standish asserts that the old regulatory paradigm of monopoly power versus regulators has somehow dissolved, leaving us with a new paradigm: integrated planning versus unacceptable risks. In the economy as a whole, pursuit of increased efficiency (and presumably profits) has achieved both objectives through ever-increasing integration both of the scale of physical goods production, and of financial and administrative corporate organization. There has been a tradeoff of efficiency versus systemic insecurity because of the vulnerability of the giant edifice to breakdown due to failure of the parts to dovetail reliably into a state of steadily increasing efficiency. In short, regulators can best serve the public interest by participating in planning with industry at both the level of states and the nation to maximize overall efficiency.

Is efficiency best served by monopoly on an ever grander scale rather than by the competition of smaller independent enterprises? His paper does not sustain the burden of proof. What does he mean by reducing the growing risks of increasing scale? Do not unacceptable risks imply unacceptable costs? And how should regulators be prepared to act against these risks? Nothing is specified.

Are we prepared to make the ideological jump which his proposal entails? The American system has presented itself as one based on competitive free enterprise. Public utility regulation was justified to the public as a means by which the aims of competitive free enterprise might be accomplished through the alchemy of legalizing and regulating monopoly enterprises. Under Standish's proposal competition drops out

entirely and the only freedom in "free enterprise" would be freedom from financial risk, underwritten by the state. How would state capitalism of this sort differ in economic essentials from state capitalism, *USSR* variety? Under the Standish model would not those with first claim on profits or surplus value be the technocratic, bureaucratic elite?

I would extend William Melody's analysis of the politico-economic aspects of the radio spectrum. There is much unrealistic talk about the alleged "scarcity" of radio frequencies. There is no general or absolute scarcity of radio frequencies, and never has been. All major radio-using countries have pre-empted and stockpiled frequency assignments which are unused or underused. In principle the radio spectrum is a limited resource at any given short period of time for certain classes of users in particular locations. Indeed the growth of spectrum use has had the form of intermittent "crises" of "congestion" as the result of successive plateaus of development, keyed to intense bursts of *R&D* for military purposes. If spectrum users and managers take certain types of action these limits are pushed back.

These actions must be considered in light of the choice between "legislated" spectrum regulation, and "positive planned" spectrum regulation. Customarily in North America, the interface between spectrum managers and licensees has been dealt with by issuing general standards and rules which legislate the conditions under which signals may be transmitted and received. This is a passive approach. Positive, planned spectrum regulation would deliberately use a variety of initiatives to push back the limits at nodes of congestion. Required *R&D* to solve specific problems, and deletion of obsolete equipment and weakly justified classes of use from certain frequency bands are examples. Another initiative involves requiring changes in the institutional structure of the users of the spectrum in order to achieve more efficient use of the spectrum. There are precedents. The *U.S.* spectrum managers required the many competitive airlines which had previously each demanded spectrum space to pool their radio services in *ARINC* half a century ago. Still another seldom used device is what engineers call "circuit discipline." This is the requirement that different entities of spectrum users integrate the use of frequencies assigned to each to increase efficiency of use (for example, through common calling frequencies)..

These types of conscious and planned action by spectrum managers are within the competence and authority of the spectrum managers and entrepreneurs. The role of economists in such an exercise should be to advise the spectrum managers and entrepreneurs, not to seek to replace them by markets or economists. In this regard, there is a rich variety of unexplored opportunities for economists to make cogent and valuable contributions to efficiency improvements.

THE DECLINE OF AMERICAN CITIES: SELF-CORRECTING AND SELF-AGGRAVATING FORCES

Population as a System in Regional Development

By WILLIAM ALONSO*

The dynamics of the population system deserve more attention than urban and regional economics has given them. This branch of economics has usually treated population as a passive variable, dependent on the demand for labor generated by changing economic forces. In the simplest case, which is also the most frequent, net migration is taken as the balancing, market-clearing instrument. Even in recent approaches which allow a greater role to the supply of labor, population is handled in a rather rigid manner: the sex-age composition is advanced by the usual Markovian techniques, with some adjustment for migration which is never theoretically satisfactory. Thus, where natural increase is considered, it is treated as exogenous; while migration is usually treated as net balances in the pure demand models, and in an *ad hoc* manner in those approaches which allow a role to the supply of population. Yet interesting issues arise when a region's population is considered as a complex system interacting with the economic system, with positive and negative feedbacks.

Consider an artificially simple example relating to net migration. An area begins to grow rapidly through net migration because of an increase in its real wage relative to the rest of the nation's. Net migration is, of course, a statistical abstraction; the real behavior consists of immigration and emigration. If the district in question is small in relation to the nation, and the number of immigrants a function of the wage difference, a constant wage difference would produce a constant stream of immigrants. But it is well established that recent immigrants into a locality have a high propensity to leave again, and that this propensity diminishes with length of residence. Consequently, places with high immigration also have high emigration. The number of those leaving would rise rapidly at first, and then more slowly as the proportion of recent arrivals diminished relative to total retained immigration. The rate of net migration, while remaining positive, would diminish over time. Or, to put it another way, in this simple model, the wage difference would have to rise over time to produce a constant rate of net immigration.

This example of a negative feedback within the population system, well documented in the economic and sociological migration literature, has been amply ignored in the literature of urban and regional economics, although migration has been the demographic process treated as endogenous in this literature. Fertility processes, which are especially important for long-run analysis, have been ignored or treated shallowly as exogenous in urban and regional economics. The balance of this paper will refer to some of the complexities of these processes.

Migration and fertility are interrelated in the growth of population and the labor force. With few exceptions (such as retirement migration) migrant streams always have the same age profile: sharply peaked in the early to mid-20's, dropping off sharply at earlier and later ages, with a smaller secondary peak of infants, who move with their young parents. Because this age profile differs from that of the regions of departure and arrival, it results typically in an older average population in the places left; and in a younger population in the places of

*Center for Population Studies, Harvard University.

arrival. In the receiving region, this migrant age composition increases the proportion of the population in the labor force, especially that of young workers, and the proportion of the dependent population consisting of children, while reducing that of older people. In the region of departure, the opposite effects are observed.

But a second effect results from the coincidence of the peak age-specific migration rates with the peak age-specific fertility rates. The crude birth rate will rise at destination and fall at origin (the crude death rate will do the opposite), in what may be termed a demographic multiplier. Disregarding now repeat migration, it can be seen that a 50-year-old migrant will contribute to his destination only the balance of his life, whereas a 20-year-old will contribute not only the remainder of his life but also that of his descendants (or, more strictly, half of these descendants, since procreation involves two parents). The effect of a wave of immigration is not smooth over time because of the years separating generations. Immigration acts upon the labor force as the coming of age of a baby boom generation, followed some twenty years later by the echo effect of the children of this generation. Again, the effects are symmetrical in the region of departure. I shall return to the parallels between immigration and a baby boom.

In demography use is made of the concept of the "reproductive value" of a migrant, developed by R. A. Fisher. This measures the number of (female) children expected to be borne by a (female) migrant of a given age, discounted by the long-term rate of natural increase. This is exactly analogous to the present value of a discounted income stream. Following N. Keyfitz, and disregarding repeat migration the reproductive value, $v(x)$, of a female migrant of age x is

$$(1)\quad v(x)=\int_x^{\beta} e^{-r(a-x)}\left[l(a)/l(x)\right]m(a)\,da$$

where β is the latest biological age of childbearing; r is the intrinsic rate of reproduction in the population (that is, the rate which would obtain in the long run if age-specific births and death rates remained unchanged) and is equivalent to the discount rate in economic models: $l(a)$ and $l(x)$ are the probabilities of a girl surviving to ages a and x; and $m(a)$ is the age-specific probability of bearing a girl at age a.

Disregarding the generational cycles, which dampen in the long run, the contribution at time t after her arrival of a female migrant of age x and her descendants to female local births is given by

$$(2)\qquad v(x)e^{rt}/k$$

where k is the mean age of childbearing in the stable population, usually about 27 years. The additional female population at that locality attributable to that migrant at time t is given by

$$(3)\qquad v(x)e^{rt}/kb$$

where b is the ratio of total yearly births to total population (i.e., the crude birthrate, but not multiplied by 1,000).

It is interesting to note that the demographic multiplier is conceptually and formally similar both to the discounting of future costs-benefits to present values and to the induced demand for labor in export-based multiplier approaches. This is because the underlying reasoning is similar, and it suggests that the formal incorporation of such an effect into urban and regional models should frequently be possible.

But more interesting than the possibility of formal links between the economic and the population systems is a major behavioral puzzle: the recent lack of convergence of regional fertility rates.

The past decades have seen a dramatic convergence of per capita income among the states and regions of the United States, and this is taken as testimony of the progressive integration of the national socioeconomy. For the same reason, it had been expected that fertility rates would converge among regions, and hence so would regional age distributions except for the effects of migration. Indeed, I. B. Taeuber and C. Taeuber, in their comprehensive study of

TABLE 1—MEANS, STANDARD DEVIATIONS, AND COEFFICIENT OF VARIATION OF PERCENTAGE POPULATION INCREASE, NET MIGRATION, AND NATURAL INCREASE, BY STATES, 1950–60, 1960–70, AND 1970–76 (DECENNIALIZED)

	1950–60	1960–70	1970–76[b]
Mean Population Increase[a]	19.7	13.4	13.5
Standard Deviation	20.1	12.6	12.2
Mean Net Migration[a]	1.2	0.8	5.2
Standard Deviation	17.0	11.1	10.0
Mean Natural Increase[a]	18.6	12.6	8.1
Standard Deviation	5.8	3.8	3.8
Coefficient of Variation	.31	.30	.47

Source: Computed from various editions of the *Statistical Abstract of the United States.*
[a]Shown in percent.
[b]Decennialized.

American demography, concluded that the convergence was so strong that in the future, geographic differentials in fertility would play a negligible role. It has been the practice in recent times, both in applied economic models such as in R. J. Olsen's et al. *MULTIREGION* and population forecasts such as those by the U.S. Bureau of the Census to postulate a rapid convergence of regional fertility rates.

However, this has not happened. Consider Table 1. The standard deviation of the states' rates of natural increase did decline from the 1950's to the 1960's, but since then it has remained constant into the 1970's. Moreover, taking into account the decline in natural increase which began in the late 1950's, we can see that instead of converging, the interstate relative differences in natural increase remained constant from the 1950's to the 1960's, and then climbed sharply in the 1970's.

These figures are not adjusted for age composition, of course, and it is therefore useful to look at the changes in the total fertility rates (*TFR*) of the various states from 1970 to 1975, which are the last dates for which data are available. The mean *TFR* declined from 2,491 in 1970 to 1,873 in 1975, a drop of 24.8 percent; but its standard deviation rose by 34.4 percent, from 218 to 293. Indeed, what happened was that, although all states declined in fertility to some degree during the period, those states which were already low in fertility had a much sharper relative drop. The regression below shows the shape of the change. Note that the coefficient of the independent variable is much above unity:

$$\ln TFR75 = -4.52 \underset{(t=14.09)}{+1.54} \ln TFR70$$
$$R^2 = .80$$

Why should regional fertility diverge when regional incomes are converging? A possible hypothesis would have this the result of two effects: a cyclical effect and a developmental effect. By cyclical effect, although this may be a misnomer, I mean the increasing or decreasing levels of local prosperity. Cyclical prosperity is generally taken in the literature to be directly correlated to fertility. In this respect, the general view is that the Sunbelt is prospering and the Frostbelt is going through hard times, which would tend to raise fertility in the former and lower it in the latter. The developmental effects, on the other hand, are viewed in the literature as a secular force leading uniformly towards lower fertility in mature economies. This is attributed to a complex of reasons which, for the purpose at hand, may be summarized as the rising direct and opportunity costs of children to prospective parents, particularly mothers. The argument is fairly well laid out for less-developed regions, such as portions of the Sunbelt, which are undergoing urbanization and sectoral and occupational transformations. The argument in the literature is

not as clearly articulated for the continued drop in fertility in the more developed areas, but it may be noted that the very low *TFR*s which the older, urbanized states have reached are very close to the current ones for comparable European nations such as the German Federal Republic, Great Britain, and Sweden.

This crude hypothesis holds, then, that in the older more-developed parts of the nation fertility is dropping both because of advancing levels of development and because of hard times. In the "newer" regions of the country, however, the effect is ambiguous, because prosperity is raising fertility while development is lowering it. The net effect would be, therefore, a sharper drop in the older, declining Frostbelt than in the Sunbelt.

Such a hypothesis is crude in part because it ignores important local circumstances. Those parts of the Sunbelt which are a modernizing and urbanizing Old South are quite different from sections further to the West, which have been highly urbanized and in a sense modern for a long time. This is true although both have a history of high fertility. Cultural and social factors are also clearly important in the case of Utah, which is prosperous and developing rapidly, but which has had both the nation's highest *TFR* and the lowest rate in its decrease.

But the hypothesis has the more fundamental problem of being based on theories developed for self-contained national populations. Urban and regional populations, by contrast, are much more open to migration. This difficulty is not just the technical one of expanding the algebra of vital rates to include the effects of migration. As illustrated above, this problem has been studied, and has essentially been solved in a variety of formulations. These approaches, however, treat the vital rates as exogenous, while the problem I am raising now is part of the determination of the rates.

Consider, as an illustration, the most prevalent cyclical theory of fertility in the United States, developed by Richard Easterlin in a series of writings (1961, 1976). Briefly stated, this theory holds that economic hardship lowered fertility during the Great Depression. After World War II, prosperous times raised fertility; fertility rose all the more because the cohort at peak fertility during the 1950's was the small cohort born during the depression which, by providing a limited supply of labor at a time of rising demand, found itself in a particularly strong economic situation. This combination of period and cohorts effects produced the baby boom. By the 1960's, however, the large cohorts of the baby boom began to arrive in the labor market, and more difficult economic times were beginning. Once again, period and cohort effects interacted in their affect on fertility, so that in the economically distressed late 1970's, twenty years after the peak of the baby boom, the total fertility rate is only half of what it was at the peak of the baby boom. By this theory, then, sometime in the late 1980's and into the 1990's, the small baby bust cohort will find itself in high demand (unless the economy is in very bad shape), and fertility will again be high.

But how does such a theory perform when regional variation in economic fortunes is taken into account? If a region is prospering relative to others, it will attract migrants, and, as mentioned above, the age composition of migrants has an effect on regional labor markets equivalent to that of a coming of age of a baby boom cohort; and to the coming of age of a baby bust cohort in regions of souring economics and net emigration. Immigration would then, by this theory, lower fertility among the original inhabitants by increasing the supply of labor provided in this fertile age, whereas emigration would be raising it in the region of departure. However, the effects of the demand for labor and its supply are operating in opposite directions, so that the net effect is unclear. More unclear yet is the size of the echo cohort, the children of those in the fertile age group, because fertility and the size of the parent group are being affected in opposite directions. How then will the supply of local entrants into the labor force be affected in a twenty-year perspective?

If a theory such as Easterlin's is employed, but complicated by the reality of

migration, it is clear that the results would vary according to the actual behavior of the various elasticities of response: migration to wage differentials, fertility to wages, wages to the supply of labor, regional development to the availability and cost of labor. One would need empirical estimates of the magnitudes of these various effects, even if one accepted the argumentation of such theorizing, to foretell their net directions and magnitudes. For econometric estimation, to say the same thing another way, one would need to use the reduced forms of a system of structural equations.

These are important effects not only for the present but also for the future course of regional development. They may be illustrated by comparing New York, a low-fertility state with emigration, with Texas, a high-fertility state with immigration. If there were to be no further interstate migration, ten years from now New York's labor force (as measured by its population between 20 and 65 years of age) would begin to decline in absolute terms, while the Texas labor force would be growing yearly by more than 1 percent. Could it be that, in a decade's time, New York will worry about labor shortages and Texas will be running with all its might to provide jobs for its growing labor force? And, of course, the current directions of interstate migration promise an earlier decline in New York and a more rapid growth in Texas.

The prospective decline of the supply of labor in states such as New York and its continued rise in states such as Texas are the product of the demographic momentum generated by their past histories of fertility and migration. They reflect cyclical effects such as those just outlined as well as the obviously important effects of development and international migration.

The purpose of this paper has been to suggest two things. First, that population, far from being an unstructured, passively dependent variable, is a vigorous one with its own internal dynamics. Second, that there are a number of interesting issues in urban and regional economics which arise from a joint consideration of the population and the economic systems.

REFERENCES

R. A. Easterlin, "The American Baby Boom in Historical Perspective," *Amer. Econ. Rev.*, Dec. 1961, *51*, 869–911.

______, "The Conflict Between Aspirations and Resources," *Pop. Develop. Rev.*, Nos. 3–4, 1976, *2*, 417–25.

R. A. Fisher, *The Genetical History of Natural Selection*, New York 1930.

N. Keyfitz, *Applied Mathematical Demography*, New York 1977.

R. J. Olsen, et al., *Multiregion: A Simulation-Forecasting Model of BEA Economic Area Population and Employment*, Oak Ridge 1977.

I. B. Taeuber and C. Taeuber, *People of the United States in the 20th Century*, a Census Monog. Washington 1971.

U.S. Bureau of the Census, "Illustrative Projections of State Populations by Age, Race, and Sex: 1975 to 2000," Series P-25, *Current Population Reports*, No. 796, 1979.

Some Dynamics of Central City-Suburban Interactions

By KATHARINE L. BRADBURY, ANTHONY DOWNS, AND KENNETH A. SMALL*

There is widespread concern about the current decline of cities in the United States, a concern encompassing such phenomena as population loss, job loss, rising unemployment rates, declining relative incomes, rising local taxes combined with declining public service levels, growing physical blight, and increasing racial and economic segregation. Without policy intervention, will such decline continue indefinitely? Or will there be a natural slowing down and eventual stop to the processes threatening the economic, social, and governmental functioning of many older large cities? Our thesis is that many of the market forces currently operating in cities cause decline to exacerbate itself in a chain of circular causation. Therefore, some policy intervention to break the cycle is in the interest of both the nation as a whole and the affected cities.

Traditionally, economists view change as an adjustment process from one equilibrium to another in response to a one-time change in conditions. For example, when transportation costs fall, people and firms need not be located as close together; the denser parts of metropolitan areas thin out, and central cities with fixed boundaries experience falling population and employment. Such a process, examined in isolation, is by its nature self-limiting: when the lower density appropriate to current transportation prices is reached, the decline stops. This stabilization operates through the land market, which balances the revised demand for space with the supply, and can take place in essentially the same manner even if continually changing conditions create a moving equilibrium. Similarly self-limiting changes are set in motion by rising average real incomes, and by new manufacturing technologies that require more space and less access to other firms and markets. Few analysts would suggest intervention to *prevent* adjustments to new situations; enlightened policy would instead be limited to easing the dislocations caused by the transition.

However, the very process of change resulting from such adjustments, or from external shocks, sets in motion other forces that may not have the same self-limiting character. The reactions of households and firms to the initial change may serve to either dampen or *amplify* the change. We call these two possibilities "self-correcting" and "self-aggravating" processes. In different contexts, others have described them as negative and positive feedback (see Vietorisz and Harrison) or cumulative causation (see Gunnar Myrdal, especially Chapters II through V). As in the example described above, "self-correcting" does not mean that the response necessarily *reverses* the change, only that it limits its extent. Self-aggravating forces, on the other hand, cause a continuous worsening after an initial negative change and therefore overshoot the new equilibrium.

Section I of the paper describes seven processes in the decline of cities relative to their suburbs that are self-aggravating in the following sense: deleterious or deteriorating conditions in a city cause rational reactions by households and firms which directly worsen the condition. Section II of the paper considers some implications. The usual justification for policy intervention in urban decline is that it imposes unfair burdens on certain members of society. If the process of change is self-aggravating, intervention is also justified to prevent overshooting. In addition, this view of decline suggests specific remedies.

*Research associate, Brookings Institution; senior fellow, Brookings Institution; and assistant professor, Princeton University, respectively. The views set forth in this article are solely our own and do not necessarily represent the opinions of the trustees, officers, or other staff members of the Brookings Institution.

I. Self-Aggravating Processes in City Decline

1) *Social discrimination*. As the "tipping" literature has amply demonstrated, if members of different groups—income classes or races, for example—are prejudiced against living or interacting (for example, through schools) with too high a concentration of other groups, an equilibrium mixture of residents (or school children) may not exist, or may be unstable. Even if some composition is acceptable to both groups, any deviation from it causes reactions that amplify the deviation. For example, whites and blacks may both value some degree of racial integration, but whites on average prefer a greater fraction of white neighbors than do blacks. If the fraction of blacks increases for some reason, the area becomes less attractive to whites and some withdraw. Their reactions reinforce the original change, further increasing the incentives for remaining whites to leave. This instability occurs even if one group concurs in the other's prejudices; for example, if both low- and high-income people prefer to have high-income neighbors. The less-favored group either "chases" the more-favored group from area to area, or, if its members are less mobile, is left behind.

In cities which have higher concentrations of black population or low-income groups than their suburbs, social discrimination at the neighborhood level leads, through net withdrawal of white or high-income population, to further increases in these concentrations at the city level. The same processes occur in public school systems. In both cases, the reactions are amplified if the first change creates expectations of a continuing process.

2) *Local taxes and local government services*. There are several related paths through which high city taxes and/or poor city services lead to further deterioration compared to suburban tax/service combinations. Because of fixed infrastructure or overhead, total costs for many local government services are not proportional to the population served. Thus if city population begins declining for some reason, per capita costs for the remaining residents rise. Similarly, if the local tax structure is more progressive than a head tax, then emigration of high-income people, for whatever reason, raises per capita taxes for all remaining residents. If in addition, the costs of providing city services such as compensatory education and public hospitals vary inversely with income, then the departure of high-income households causes *costs* per capita to rise, and *taxes* per capita to rise even more. Finally, selective departure of upper- and middle-income groups erodes the political power of such groups in the city government, which can cause a shift toward public activities particularly valued by the poor. All these reactions make the city less attractive to high-income people in comparison with a more homogeneous suburban jurisdiction, and all except the last harm the poor residents as well. Barriers to low-income entry—so-called "fiscal zoning" measures such as minimum lot size requirements and high-quality construction codes—prevent the poor from sharing in suburban gains.

Similar forces operate on the location decisions of firms, which also pay local taxes and consume local government services. In both cases, a fall in the economic base causes tax increases or service declines which induce a further emigration of those jobs or residents contributing to the city's ability to finance services.

3) *Agglomeration economies among firms*. Cities evolved largely because firms found it profitable to operate near input suppliers, markets (firms and consumers), or other firms engaged in similar activities. Such "agglomeration economies" accrue to firms in proportion to the mass of other relevant firms in the vicinity, at least until congestion becomes a problem. If a city begins to lose some firms, each departure reduces the opportunities for interaction and thereby the profitability and attractiveness of the city for the remaining firms.

4) *Critical mass activities*. Some city services and amenities require a high volume of users to operate efficiently. These decreasing-cost activities include mass transit services, diversity in restaurants, specialized retail outlets, and cultural activities. If

the patronage of such institutions declines, average costs rise. This forces price increases and/or service cutbacks, or else calls forth a public subsidy, which feeds back through the fiscal system (unless it is externally financed). In either case, the decline is reinforced.

5) *Physical blight and abandonment*. The value of a housing unit on the market and to its current occupants is affected by the condition of other housing units in the immediate neighborhood. If one unit is severely damaged by an accident or left vacant for an extended period because of net population losses in the area, it detracts from the desirability of the entire neighborhood. The subsequent demand reduction in turn reduces the profitability of landlords' maintenance efforts. Additional units will run down, and may even be abandoned, and the demand for neighborhood units falls further in response. In this way, physical blight and abandonment are contagious through a process of self-fulfilling expectations. While property values fall to reflect these demand shifts, no one owner can guarantee what will happen to nearby units. Hence the recovery of the area requires a reversal of each owner's negative expectations about the behavior of others.

6) *Crime*. Cities generally have higher crime rates than their suburbs. These higher rates induce those who wish a more secure environment to locate in the suburbs or to stay indoors. This increases the incidence of crime among the remaining residents both because streets are more deserted and because those who commit crimes become more spatially concentrated.

7) *Political power*. As cities lose population, they lose political power within the metropolitan area and in the state and national arena. As we discuss in the next section, the most useful policies for fighting cycles of city decline include several that must come from outside the city. Thus, just as it needs more assistance, a city's political power in higher-level governments' decision making wanes. Failure to obtain assistance leads to further population—and power—losses.

Although we have discussed these seven processes separately, they can operate simultaneously, and their interactions further exacerbate decline. For example, as a neighborhood begins to show signs of serious physical decay, those residents most able to move do so; this isolates the poor and unemployed, and may lead to increasing crime rates. Vandalism augments the physical decay, and retail stores or other local employment sites may become unprofitable. Their departures contribute to additional unemployment and the cycle continues. A second example of interaction is the possibility that the income segregation tendencies described in 1) would set off the fiscal decline cycle described in 2).

II. Implications: Maintaining Responsibility for Social Costs

Certain restraints on the processes described above usually act more or less automatically. For some processes, institutional structures and relations created by the opposite process tend to retard decline. For example, businesses concentrated together taking advantage of agglomerative economies have established institutions, channels of communication, and personal relationships which continue even after out-movement begins, and hence tend to retard it. Similarly, an entrenched political organization may retain power disproportionate to the size of its constituency. These automatic restraints are not sufficient to halt adverse change entirely; but public policy aimed at preserving such ties and institutions may be a relatively cheap way of reducing the self-aggravating nature of decline.

For those processes involving population losses, a restraining factor might be the desirable aspects of lower residential density and less congestion. However, their positive effects are limited by the particular patterns in which population and job losses occur. On the job side, it is the congested downtown area which is least affected by decline. On the population side, abandonment of buildings or apartment units results in an uneven "thinning out," and in side effects

such as arson, drug traffic, and havens for derelicts. Living in a block of 100 half-abandoned row houses is quite different from living in a fully occupied block of 50 detached houses!

Many of the processes described in the previous section clearly involve elements of increasing returns to scale or externality, which suggest classical economic remedies. The remedy of tax-subsidy policies is currently utilized through a variety of programs which attempt to influence the location of firms, to subsidize critical mass activities, and to encourage housing maintenance and renovation. The remedy of internalizing the externality is exemplified by redevelopment authorities that coordinate the reconstruction or renovation of large areas within a city. Such programs can be quite cost effective to the extent that they reduce those deleterious changes which are magnified by the processes described here. A third classical remedy is to facilitate side payments—to create a market for the externality—in this case, to bribe movers to remain. This appears to be impractical in the current context.

The problem with the kinds of corrective forces or actions described thus far is that they act primarily to achieve static efficiency. If population density is too high, people will continue to leave until the newer densities are satisfactory. If the level of building maintenance is too low, subsidies or social pressure may bring it up to a desirable level. But the distinguishing feature of the processes described in the previous section is their *dynamic* nature, two key features of which are prominent: *mobility* and *expectations*. Each of the processes involves emigration and thus works by taking advantage of geographical mobility. At the same time, since such moves entail market transactions for land or long-lived capital whose prices are sensitive to expectations about future conditions, the strength of the self-aggravating forces depends on the expectations held by individuals. Indeed, one household's decision to move may depend on *its* expectations of the *future* expectations of potential neighbors whose identity is not even known! To the extent that expectations are based on small *changes* in current conditions, these processes display a volatility which distinguishes them from most of the "classical" market failures studied by economists.

As is well known from the study of feedback systems in engineering, one cure for excessive volatility is to introduce resistance or friction into the system. In the present context, this implies policies to *reduce* mobility; for example, a requirement that city government employees live inside the city. This is in direct conflict with the policies needed to cope with transition from one static situation to another, and could be seriously recommended only if the dynamic distortions were felt to be a greater problem than slowness to adjust to changed circumstances.

The problem is really one of *exit*, in the sense used by Albert Hirschman. By withdrawing from an entire neighborhood or community, an individual or firm is able to avoid its problems, but in so doing aggravates them for those who remain. As Hirschman noted in a different context, the existence of too easy an exit option may lead to results which are far from optimal for society as a whole.

One way to discourage exit, aside from lowering mobility, is to enlarge the relevant boundaries: by merging political units, by mixing school populations, or through power-equalizing arrangements (which enlarge the area within which the tax price of public services is uniform). More broadly, exit is discouraged if it does not entail release from responsibility for the unfavorable condition. For example, state and federal aid to cities make *all* taxpayers at least partially responsible for meeting city fiscal needs, even if they live outside city boundaries. For the case of social discrimination, in which differential mobility allows one group to exit and thereby escape contact with another, policies which enhance the mobility of the less-favored group may cause members of the more-favored group to be discouraged from leaving by the expectation that they will be followed. Such

policies include more rigorous enforcement of laws against racial discrimination in housing, raising the incomes of the poor so they can enter higher-cost areas, creating subsidized housing in such areas, and reducing barriers that block movement. Examples of the last are building codes and zoning regulations that deliberately raise housing costs far beyond levels required by health and safety.

Another way to discourage exit is to offer a desirable alternative means of changing conditions: in Hirschman's terminology, to facilitate *voice* in urban affairs. Thus, a responsive city government may be able to alleviate some conditions which result more from inefficiencies in operation of public services than from any underlying necessity. The encouragement of civic pride may cause some individuals or firms to remain and work to improve their city because they are given an avenue to convince others to do the same. Similarly, neighborhood organizations can provide a means for otherwise isolated individuals to agree to desist from behavior which aggravates neighborhood change, thereby stabilizing mutual expectations about the neighborhood's future. For example, some neighborhoods in the eastern suburbs of Cleveland have adopted an affirmative policy of racial integration to prevent tipping to either all black or all white. Citizen crime patrols and neighborhood "whistle stop" campaigns play a similar role in thwarting crime and the fears it creates.

All these policies attempt to break the chain of cumulative causation. In some cases, it is more feasible to reverse the process than to break some link in the cycle. For example, large-scale public or private investments in declining downtowns, such as Detroit's Renaissance Center, provide a new focus for the growth of agglomeration economies. In these cases, a policy push attempts to make the cumulative decline process work in the opposite direction.

III. Conclusion

The problems of central cities are viewed by many as simply transitional difficulties in adjusting to changed economic circumstances. From this point of view, public policy should facilitate adjustment on the part of firms and individuals: in short, allow them to exit gracefully.

But we believe many forces set in motion by such transitions do not contribute to the needed adjustments, but instead lead to inefficient and costly moves designed to escape actual or anticipated deterioration in conditions. The existence of these forces suggests the need for a quite different set of policies to break into the positive feedback cycle. These include enlarging the relevant boundaries either literally or figuratively to make escape from responsibility more difficult, reducing barriers to the mobility of groups now being left behind, and providing for "voice" as an alternative means of both influencing conditions and stabilizing expectations. Less frequently, policies may be devised to reverse the cumulative causation process to the city's advantage. To do nothing is to assume that market forces will bring about needed adjustments efficiently, a doubtful presumption given the complex interplay of forces involved in city decline.

REFERENCES

Albert O. Hirschman, *Exit, Voice, and Loyalty*, Cambridge, Mass. 1970.

Gunnar Myrdal, *Rich Lands and Poor*, New York 1957.

T. Vietorisz and B. Harrison, "Labor Market Segmentation: Positive Feedback and Divergent Development," *Amer. Econ. Rev. Proc.*, May 1973, *63*, 366–76.

Neighborhood Stability in Changing Cities

By MICHAEL A. STEGMAN AND DAVID W. RASMUSSEN*

No matter how defined, the neighborhood plays a significant role in the life experiences and well-being of the urban population. This fact is consistently demonstrated empirically in mobility studies (see J. H. Mark, T. P. Boehm, and C. L. Leven) and in multivariate estimates of the determinants of housing prices (see J. R. Follain and S. Malpezzi). Hedonic pricing models confirm the proposition that locational and related neighborhood attributes are valued components of the housing bundle. This paper addresses the determinants of neighborhood stability within a dynamic market setting. In this introductory section we outline the extent of recent changes in the housing stock of older central cities. The role of neighborhoods in models of the housing market is then explored in order to define neighborhood stability in the context of normal market processes. The paper concludes with a discussion of neighborhood instability and public policy.

Other things equal, housing quality varies inversely with the age of cities and with the age of housing stock (see Richard Muth, 1978, p. 76). These independent effects of age exacerbate the housing market impacts of a decline in the socioeconomic status of the population with the result that downward stock adjustments in older declining cities can be substantial.

While no national data document the relative importance of the various means by which housing suppliers reduce the size of their inventories during periods of market distress, evidence suggests that where local market and neighborhood conditions permit, relatively large numbers of rental units are sold to owner-occupants. Thus, for example, between 1970 and 1976, the percent of single family homes in central cities occupied by renters declined from 20.2 to 17.3 percent or by close to 800,000 units (see U.S. Bureau of the Census, Table A-1). On the multifamily side, condominium conversions are also on the rise, with the number of converted units estimated to reach 130,000 in 1979 alone (see K. Harney). Conversion is, of course, a supply-side response that is reversible over the longer term. Even today, it is estimated that as many as 15 percent of all converted condominium units in Chicago were acquired for speculative purposes by non-owner-occupants and remain part of the multifamily rental sector (see J. Wilmann).

With respect to more permanent responses to market decline, especially where the downturn is expected to be long term or in certain neighborhoods where deterioration is already quite substantial, abandonment is a (dis-)investment strategy being elected with increasing frequency. Between 1973 and 1976, net of conversions to owner occupancy, the number of rental units removed from the nation's inventory equaled 4.3 percent of the 1976 rental stock (see U.S. Bureau of the Census, Tables A-1, A-5). Reasons for removal are not limited to abandonment, although data for New York City, for example, indicate that such investor actions are the principal cause of gross losses from the stock (see P. Marcuse, p. 76).

Ignoring for the moment the issue of whether the growing incidence of demolition and abandonment reflects suboptimal resource use, a recent General Accounting Office (GAO) report indicated that 113 major cities have experienced the abandonment phenomenon, at least to some degree (see GAO, 1978b, p. 8). As of mid-1973, in excess of 4 percent of Philadelphia's housing

*Deputy Assistant Secretary for Research, U.S. Department of Housing and Urban Development, on leave from the department of city and regional planning, University of North Carolina-Chapel Hill, and professor of economics, Florida State University, respectively. The views expressed in the paper are our own and do not necessarily reflect the official views and policies expressed or implied of the U.S. Department of Housing and Urban Development.

stock was abandoned, as was almost 3 percent of St. Louis' residential inventory (see GAO, 1978b, p. 14). New York's tax foreclosed stock now exceeds 11,700 multifamily buildings, including 4,100 which are still occupied (see City of New York, 1979, p. 10). About 45,000 privately owned properties in the city are now in sufficient tax arrears to be eligible for foreclosure: Unless in rem actions are either discontinued or slowed down, New York's ownership share of the multifamily stock is sure to increase over time (see City of New York, 1979, p. 6).

Indications of other supplier adjustments to market decline can be gleaned from isolated data for individual cities. In Chicago, for example, there has been a substantial decline in investor purchases of rental-housing tax liens at city auctions (see J. J. Lawlor, p. 2). One explanation for this slowdown is that tax delinquent owners are increasingly reluctant to redeem their properties. Lien purchasers deal in interest-bearing paper, not tenants. When landlords fail to redeem their properties in any significant number, the supply of bidders dries up. While no comparable redemption data are available for New York City where a tax lien is satisfied through auction of the delinquent property itself, a recent study indicated that more than 90 percent of auctioned properties were in default of their taxes (for city-financed purchase money mortgages) within two years after sale (see City of New York, 1978).

Two irreversible investment strategies are alleged to be occurring with some frequency in declining markets. Welfare loading is a practice of filling a building with public assistance families at maximum permissable rents while providing minimum levels of tenant services. Often, welfare loading precedes abandonment and is simply a way of withdrawing sunk capital from the physical stock. Arson for profit is another means of extracting sunk capital from buildings; only in this case the return is received in the form of a one-time insurance payment. Estimated arson property losses in the United States exceeded $630 million in 1975, an 829 percent increase since 1964 (see GAO, 1978a, p. 6; J. Newfield and P. DuBrul, p. 5). Even more recently the GAO reported that in many cities across the country, so-called FAIR plan operations, which are supposed to make available fire and hazard insurance to properties in older neighborhoods, are creating incentives for arson for profit.

Substantial stock adjustments have major impacts on neighborhoods. The concept of stable neighborhoods can only be discussed in the context of this rapidly changing urban environment. Because of market interdependencies, and the fact that the disinvestment process is neither systematic nor orderly, larger numbers of housing units may be adversely affected by a given market decline than would be warranted solely on the basis of lower demand. The full stock-adjustment process, which includes second- and subsequent-round supplier responses to isolated instances of abandonment, etc., occurs at the neighborhood scale where externalities abound. It is at the neighborhood scale, therefore, that the issue of self-correcting or self-aggravating market forces must first be treated.

I. The Role of Neighborhoods in Urban Models

Early models of housing market behavior assumed housing to be a homogeneous good that could be described by monthly rent or housing value (see Edgar Olsen; R. Muth). This approach was used to develop a competitive theory of housing markets and allowed precise definition of commonly used housing terminology. For example, within the competitive model, filtering is defined as a change in the number of housing services in a particular dwelling unit. Since the price per unit of housing services always moves toward equilibrium, the impetus for filtering rests on the existence of price differentials in submarkets offering varying units of housing service per dwelling.

As the basic dynamic force in the competitive market model, filtering is an essentially aspatial and symmetrical process. Rather than geographically defined, housing submarkets are differentiated in terms of service flows, while the economics of filtering up and down are basically the same.

When strictly construed, therefore, the competitive model cannot readily explain neighborhood instability in terms of persistent differences in relative prices among city neighborhoods since price differences in a competitive market are a short-run phenomenon. It follows, therefore, that neighborhood or submarket decline must be defined in terms of quality changes or the downward filtering of classes of dwelling units; a normal market phenomenon we do not necessarily wish to equate with the kinds of conditions and irreversible investor actions characteristic of seriously deteriorating neighborhoods.

The flexibility of the competitive model is reflected in its recent extensions to incorporate neighborhood considerations into market-pricing processes. G. K. Ingram and J. F. Kain, and B. Bender among others demonstrate that neighborhood attributes affect dwelling values. They postulate that the unit price of housing services is invariant within, but variable across, neighborhoods. Within a neighborhood, individual investors maximize the net present values of their dwellings by adjusting the level of housings services to the prevailing neighborhood price structure. Since the stock price reflects the net present value of an income stream, the investment approach to modeling market activities emphasizes the costs of supplying housing services, investor's expectations, and residual land values, as well as current revenues.

The expansion of the competitive model to include neighborhood effects notwithstanding, market equilibrium at the neighborhood level and neighborhood instability are difficult to define. Market prices, certain variable and fixed costs, real depreciation rates, and the number of buyers willing to bid on available properties may all vary considerably by neighborhood (see Michael Stegman, Ch. 3). Within this dynamic process, a neighborhood equilibrium may be said to exist when dwellings are not being filtered or when an area's dwellings are receiving the best possible maintenance consistent with normal profits. In popular terminology, these are stable neighborhoods. A neighborhood in disequilibrium, one whose housing stock is being filtered in response to exogenous changes in conditions of supply or demand, may be stable so long as the market adjustment process ultimately produces a normal rate of return to investors. In short, a neighborhood housing market is stable when it is in equilibrium, or, if in disequilibrium, when there exists an equilibrium solution. By equilibrium solution is meant the existence of some achievable combination of dwelling values, rents, vacancy rates, and operating costs that will result in normal returns to neighborhood investors. Neighborhood instability exists when there are no achievable operating and market conditions which will yield normal profits. Self-correcting market forces are operative where neighborhood equilibria exist or are possible and are inoperative or self-aggravating where equilibrium conditions do not or cannot exist.

II. Neighborhood Stability and Public Policy

Individual and collective supplier responses to substantial market decline may result in the scrapping and/or abandonment of neighborhood housing stocks. Following Ingram and Kain, a housing unit is scrapped when it produces a lower service flow than is permitted by law or accepted by the market. Absent neighborhood effects, scrapping may be consistent with efficient market operations, as newer, more desirable dwellings in more modern and spacious surroundings drive older and least preferred ones from the inventory. However, because significant externalities are an inherent part of the stock-adjustment process, older neighborhoods whose housing stocks would otherwise adjust systematically to the marketwide decline in housing demand may become unstable. It is in these neighborhoods that the widespread incidence of scrapping and abandonment is inconsistent with the optimal use of housing resources.

Public policies designed to reduce neighborhood instability must recognize that publicly supported efforts to expand home ownership rates and the metropolitan housing supply frequently result in suburban growth and a corresponding decline in the

demand for central city housing. While there is little empirical analysis that causally links new construction rates on the urban fringe to the widespread scrapping of central city housing, the expected high correlation of the two is sufficiently great for Anthony Downs to argue that "the main cause of abandonment of central cities is the construction of suburbs" (p. 25).

Declining neighborhoods would appear to be unavoidable in cities that are aging, losing households, and where substitute new housing is being built in the suburbs. No neighborhood stabilization policy can ignore this reality. Three policy perspectives are suggested by this analysis. First is *laissez-faire*, a policy with considerable appeal to many economists. Predicated on the neoclassical assumption of perfect mobility, this view implicity discounts the utility of neighborhood ties in housing decisions and market operations. Negative externalities are recognized at the neighborhood level but are viewed as less costly and less disruptive to the market than any conceivable program that might be implemented to reduce their deleterious effects. A second perspective recognizes the inevitable substitution of newer housing for older, but seeks to minimize that portion of disinvestment that is caused by externalities rather than by reduced demand per se. Policies consistent with this view, frequently referred to as planned shrinkage, managed migration, creative thinning and triage, emphasize concentrating scarce program resources in the most viable neighborhoods (see W. C. Baer). The implementation of neighborhood stabilization programs on a selective basis effectively targets the relatively few locations in the city in which effective housing demand will be shored up with the assistance of the public sector. By implication, nontargeted neighborhoods will undergo accelerated decline, thus allowing the land to be "banked" for the eventual reconstruction of the central city. The third perspective relies more heavily on the psychological and sociological dimensions of household housing decisions and takes explicit note of neighborhood considerations in personal well-being. By implication, this view emphasizes the utility households derive from neighborhood ties and the possibility of nonmarket explanations for market behavior. Within this policy context, no neighborhood should be "written off" on efficiency grounds because urban neighborhoods serve useful social purposes that cannot be measured by price and related market data. Similarly, this view suggests that stabilization programs will not work in every neighborhood since variables like community cohesiveness and the effectiveness of local leadership substantially influence household and investor housing decisions. This perspective suggests a neighborhood-based policy that makes all areas eligible for assistance, with the specific program treatment implemented depending upon such variables as the effectiveness of neighborhood groups in the reinvestment process, the extent of neighborhood decline that has already occurred, etc.

III. Conclusions

The presence of self-correcting and self-aggravating forces at the neighborhood level is partly a function of time horizon. In the long run, as economists since Keynes are fond of pointing out, aging defeats people and material structures. Since everything is variable in the long run, the fact that the market may be self-correcting at the neighborhood level for housing assets which last over half a century is not very interesting. In the shorter run, given the immobility of the housing stock and the interdependencies that characterize urban neighborhoods, self-correcting market forces are not always operative. In this paper we have defined neighborhood stability in a dynamic way that permits the normal aging process to operate—a neighborhood is stable if normal profits can be earned after a period of market adjustment. For neighborhoods in the poorest condition, particularly those with no comparative site advantages, however, there may not be any self-correcting market forces which can halt the decay.

REFERENCES

W. C. Baer, "On the Death of Cities," *Publ. Interest*, No. 45, Fall 1976, 13–19.

B. Bender, "The Determinants of Housing Demolition and Abandonment," *Southern Econ. J.*, July 1979, *46*, 131–44.

A. Downs, "Government Organization and Urban Economic Development," in *Urban Economic Development: Past Lessons and Future Requirements*, House Committee on Banking Finance and Urban Affairs, Subcommittee on the City, Washington, July 1979.

J. R. Follain and S. Malpezzi, "Dissecting Housing Value and Rent: Estimates of Hedonic Indexes for Thirty-Nine Large SMSAs," unpublished manuscript, Urban Institute 1979.

K. Harney, "10% of Rental Conversions This Year Will be in D. C.," *Washington Post*, Apr. 28, 1979.

G. K. Ingram and J. F. Kain, "A Simple Model of Housing Production and the Abandonment Problem," *Amer. Real Estate Urban Econ. Assn. J.*, June 1973, *1*, 79–106.

J. J. Lawlor, "Real Property Tax Delinquency and the Rehabilitation of the Multi-Family Housing Stock in Chicago, Illinois: The Role of the Collection Provisions of the Illinois Revenue Act," *DePaul Law Rev.*, No. 1, 1976, *26*, 1–22.

N. Levanthal, "Analysis of In-Rem Workload in New York City," mimeo., City of New York, Feb 28, 1979.

P. Marcuse, *Rental Housing in the City of New York: Supply and Condition 1975-1978*, New York, Jan. 1979.

J. H. Mark, T. P. Boehm, and C. L. Leven, "A Probability Model for Analyzing Inter-neighborhood Mobility," in D. Segal, ed., *The Economics of the Neighborhood*, New York 1979.

Richard Muth, *Cities and Housing*, Chicago 1969.

______, "Regional Differences in Central-City Housing Quality," in *Urban Housing: Proc. Fourth Annual Conference, Fed. Home Loan Bank San Francisco*, 1978, 73–88.

J. Newfield and P. DuBrul, *The Abuse of Power: The Permanent Government and the Fall of New York*, New York 1977.

E. O. Olsen, "A Competitive Theory of the Housing Market," *Amer. Econ. Rev.*, Sept. 1969, *59*, 612–22.

Michael A. Stegman, *Housing Investment in the Inner City: The Dynamics of Decline*, Cambridge, Mass. 1972.

J. Wilmann, "Are Condo Investors Really Aiding D.C. Renters?," *Washington Post*, Nov. 11, 1979.

City of New York, "The In Rem Housing Program, First Annual Report," Oct. 1979.

______, "In Rem Pricing Study," Department of Housing Preservation and Development, Nov. 1978.

General Accounting Office, (1978a) *Arson for Profit: More Could Be Done to Reduce It*, Washington 1978.

______, (1978b) *Housing Abandonment: A National Problem Needing New Approaches*, Washington 1978.

U.S. Bureau of the Census, "Annual Housing Survey: 1976," Part A, Washington 1978.

DISCUSSION

HARVEY A. GARN, The Urban Institute: William Alonso notes that urban and regional models have generally assumed that employment growth influences population growth with population migration treated as the mechanism which equates labor supply to demand. There has been recent discussion in the literature which suggests that the influence of population growth on employment has been growing and may be the major influence. Those who argue this case attribute the change in strength to increasing importance of markets in industrial location partly as a result of shifts in the overall distribution of employment toward services and away from manufacturing. My reading of the evidence is that, even though both effects are clearly operating, the predominant effect is still a population response to patterns of employment change rather than the reverse. The issue is of considerable importance from a policy perspective, since it influences the mix of programs which might be suggested for stemming the joint decline of employment and population from many of our large urban areas. The question is whether one would expect better results from a mix which puts most weight on direct efforts to influence the location and expansion of business firms or on efforts to influence the location of households through improvements in residential areas and services.

It remains true, however, that the specific requirements for (let alone the long-run desirability of) inducing migration or changes in natural population increase (births over deaths of the nonmigrant population) in cities which are losing population and jobs have not been determined. In illustrating the feedback effects of migration on fertility due to changes in the age distribution, for example, Alonso starts the discussion by assuming a migration response to real wage differences between places and argues that an increasing real wage difference is necessary to induce a constant net flow of migrants over time. This conclusion follows from the assumed migration response to area-wide real-wage differences.

Recent studies by Morrison and others have shown, however, that migration in the 1970's is not explained by such an assumption. There is considerable migration from higher to lower income places, resulting in substantial net population gains for such lower income areas. A set of work force migration studies which we have been conducting at the Urban Institute has shown that both interregional and intermetropolitan sector migrants generally experience greater increases in mean wages than nonmigrants regardless of the relative income levels of origins and destinations. This suggests that general wage differences between areas may not be as important in determining migration flows as the wage differences expected by individual migrants. These latter expectations are not necessarily correlated with area wage differences. Other recent migration studies (see, for example, Ann Bartel's 1979 paper in the *American Economic Review*) have shown that migration probabilities are influenced by job tenure, quitting a job, and being separated from a job which are not uniformly correlated with wages. It is not necessarily the case, therefore, that wage differences between areas would have to rise over time to produce a constant rate of net immigration.

Alonso's major point, however, is that migration produces further endogenous changes in the population of both the sending and receiving areas. Migration tends to make net receiving areas younger and losing areas older. These age-distribution changes create further changes in the total fertility rates in both kinds of places. Further endogenous change in both migration patterns and fertility rates of nonmigrants might be produced by changes in job availability, stability, wage rates, and externalities attributable to the location decisions of migrants.

Repeat migration is an interesting part of the puzzle. Bartel suggests that this has to do with job tenure and the decision to change jobs, not simply length of residence. Our migration studies show that high-immigration rates are generally associated with high outmigration rates for regional

and metropolitan scale aggregations. The importance of return migration, in the context of Alonso's paper, is that it is risky to predict the long-term addition of descendants of recent migrants (a delayed baby boom) in the places which are now experiencing the highest rates of migration. Many of these, too, will pass to other places.

Diverging fertility rates may be partly explained by aspects of the selectivity of migration in addition to migrants' systematic age profiles as well as endogenous changes in fertility. Alonso's paper, however, does show that an array of factors linking the population and economic systems in ways that are little understood could exert an endogenous influence on fertility as well as migration. While the paper does not establish that such changes in fertility are likely to overwhelm the direct effects of migration on job and population change in subnational economies, it does support his case that the joint interaction of the economic and population systems have received too little attention in urban and regional economics.

WILLIAM H. FREY, Center for Demography and Ecology, University of Wisconsin-Madison: I agree with Katherine Bradbury, Anthony Downs, and Kenneth Small (hereafter, B-D-S) that there has been in the past, and continues to exist, a complex of self-aggravating mechanisms in our most distressed central cities. However, given the present condition of these cities and what is known about their population dynamics, I believe that the feedback aspects of the system may be less devastating than B-D-S suggest; and that the substantial interventionist policies required to ease the "aggravation" will likely not make all that much difference for the eventual equilibrium state reached.

First of all, the significance of the city-suburb redistribution dynamic appears to be becoming a less-important source of city decline. The distressed cities we are discussing are highly dense and severely underbounded jurisdictions whose housing and population compositions stand in sharp contrast to those in the remainder of the metropolitan area. As a result they are disproportionately composed of elderly, poor, and black households that are far less prone to mobility than the majority of metropolitan residents. Indeed a more important city problem than the increased out-movement of its upper-income residents is the immobility of this residual population which has been left behind by waves of postwar suburbanization. While it is true that the city-suburb movement stream still constitutes the greatest demographic component of city loss, data I have compiled show the magnitude and selectivity of this stream to be decreasing over time. It appears that the increased losses older cities are now sustaining can be attributed to lower rates of natural increase and greater net out-movement from the metropolitan area—two processes that are not directly linked to the city-suburb redistribution dynamic.

The second point is that, even if achieved, the altered city-suburb conditions proposed by B-D-S are not likely to effect substantial changes in aggregate movement patterns in the short or moderately long term. While it has been said that people "vote by moving" to maximize consumption of residential services, they vote neither early nor often in response to changing community conditions; and when they do their response is not as rational as economists might like to think. The individual mobility process is laden with inertia, and changing community characteristics are less likely to affect the decision to move—typically a response to life cycle transitions—than they are to influence the destination choice of individuals who have already decided to move. Empirical evidence suggests that even this choice will be determined less by the community conditions B-D-S would influence, than by the availability of a specific dwelling unit type (for example, single-family dwelling unit of desired size).

Finally, I want to discuss some implications of life cycle related mobility patterns which are relevant to the B-D-S thesis. While it is well known that the frequency of movement is concentrated within one's young adult years (ages 18–35), empirical studies have also shown a life cycle related-

ness to movers' city and suburb destination locations. During their late teens and early 20's, individuals have shown a tendency to locate in cities—largely because the centrally located, rented housing available there proved compatible with the needs of single, group quarters and childless-couple living arrangements they entered into. As these same individuals aged to their late 20's and 30's, suburban locations became more attractive—again as a result of the housing availability (single dwelling units in low-density neighborhoods) which was consistent with the married-couple-with-children living arrangements that predominated in these ages. The point here is that past population cohorts were already located in the suburbs as they entered their nonmobile years, and it was difficult for central cities at that point to draw them back.

Recent changes in population living arrangements—toward smaller households and fewer children—suggest that past patterns need not be repeated. If cities are able to provide housing attractive to the now smaller young adult households, they may not only be able to attract them during the late teens as they had in the past, but also retain them during their less mobile years, as a result of a more attractive housing stock. Should this occur—and it certainly will not happen overnight—then the measures proposed by B-D-S may appreciably affect the pace of city population change. However, operating only by themselves, these measures are not likely to keep city family-forming households from leaving for more attractive housing and neighborhood conditions in the suburbs; nor will these measures attract back to the city, suburban families that have already reached the more stable (nonmobile) life cycle stages.

JANET ROTHENBERG PACK, School of Public and Urban Policy, University of Pennsylvania: The paper by Michael Stegman and David Rasmussen (hereafter, S-R), and the paper by Katherine Bradbury, Anthony Downs, and Kenneth Small (hereafter, B-D-S), point to urban phenomena and scenarios we can all recognize in broad scope. However, they suggest questions for which there are currently few answers. For example, how much additional abandonment of structures results in an abandoned neighborhood as a result of an initial abandonment, compared with what would have resulted from the underlying conditions, such as the aging of the structures, their increased costs of maintenance, declining demand? How much additional movement to the suburbs results from the increase in the tax burden caused by the previous movement of upper-income households that would not have occurred as a result of rising tax burdens associated with increased public expenditure? Indeed, is this general line of argument consistent with the literature that finds tax and expenditure differences to be capitalized in house values? It may even be that the enormous expansion of federal government aid to local governments has blunted the force of this argument. Between 1959 and 1977, federal aid to states and localities increased from $2.3 to $70.4 billion, and from about 10 to 27 percent of state and local expenditures. The figures for central cities would be substantially greater. Moreover these funds are concentrated in precisely those areas cited as increasing the costs of center city governments: public welfare; health and hospitals; housing and urban renewal; and education.

Still other questions are suggested by differences among neighborhoods and cities. Is the existence and magnitude of negative externalities and positive feedback invariant to context? Will an abandoned structure on a block with single-family owner-occupied structures have the same implications as it would on a block with multidwelling unit structures with nonresident owners? Does the loss of a firm or of an entire industry always result in cumulative loss without replacement, irrespective of the degree of specialization of the industrial structure of the city, of the public services it provides, or the skills of its labor force? Are the assumptions of continuous positive feedback consistent with the alleged movement back to the center city (or decreased out-movement) by

middle- and upper-income households, observed not only in cities we think of as special, such as Washington D.C., and San Francisco, but also in Philadelphia, Boston, and Baltimore?

In reading both these papers, the Forrester model (Urban Dynamics) came to mind again and again. While there are very important differences, the similarities are striking. Among these are: 1) the emphasis on negative externalities, positive feedback, and the inevitability of decline; 2) the importance of the aging of structures is central to S-R and to Forrester; in the latter, it is the increased occupancy of urban space by older structures that leads to urban decline; 3) high tax burdens on upper-income groups, associated with disproportionate expenditure on low-income groups, exacerbates urban decline in both B-D-S and Forrester; 4) in all three, the decreasing relative attractiveness of the urban area relative to other areas causes middle- and upper-income families to relocate and thus contributes to further urban decline.

I am not suggesting that we turn to the Forrester model for guidance for reasons that have been amply documented in many places. However, serious evaluation of the issues posed would seem to require a fairly comprehensive modeling effort. Surely the B-D-S seven self-aggravating processes are interrelated and could not be tested seriatim. Similarly in S-R it is not simply the spillovers from dwelling to neighborhood and back to dwellings which needs to be explained, but also the spillovers among neighborhoods. Models are required not only to verify the direction of association posited but also their quantitative importance.

We need to articulate more precisely and verify these hypotheses if policy recommendations are to be evaluated. In B-D-S consolidation of metropolitan area jurisdictions would make the tax base the metropolitan tax base; industry and upper-income households could not escape their responsibilities by moving to the suburbs. What would the tax and expenditure packages look like in the consolidated jurisdiction? Would there be a rearrangement among metropolitan areas over time just as there was a redistribution of population and industry from city to suburbs? Would the merger of city and suburb have any effect on the unstable neighborhoods with which S-R are concerned? Given fixed budgets could either neighborhood investment strategy prevent declining neighborhoods from becoming unstable or turn unstable ones around? How would other areas of the city be affected?

Given the importance of the issues raised in these papers, the persistence of the problems cited despite substantial policy initiatives, and the calls for still further policy intervention, it may be worthwhile to continue to develop comprehensive modeling efforts despite their substantial inadequacies. Such efforts require the sustained support of more detailed partial analyses. The past shortcomings of the comprehensive models and the more successful development of partial analysis may have widened the gap between the two.

AMERICAN ECONOMIC ASSOCIATION

PROCEEDINGS

OF THE

NINETY-SECOND

ANNUAL

MEETING

ATLANTA, GEORGIA

DECEMBER 28–30, 1979

THE JOHN BATES CLARK AWARD

Citation on the Occasion of the Presentation of the Medal to

JOSEPH E. STIGLITZ

December 29, 1979

The terms of the John Bates Clark medal call for its award every second year "to that American economist under the age of forty who is adjudged to have made a significant contribution to economic thought and knowledge." That would seem to rule out Joseph Stiglitz, except that we interpret the phrase "a significant contribution" to signify a distant lower bound. Stiglitz is beyond compare among younger economists for the range and variety of his theoretical achievements, as well as for their vigor and their liveliness. From growth and capital to the economics of discrimination, from public finance to corporate finance, from information to uncertainty, from competitive equilibrium with exhaustible resources to monopolistic competition and product diversity, contemporary economic theory is crisscrossed with his footprints. In confidence that the trail still has a long long way to wind, we award the John Bates Clark medal to Joseph E. Stiglitz.

Minutes of the Annual Meeting
Atlanta, Georgia
December 29, 1979

The Ninety-Second Annual Meeting of the American Economic Association was called to order by President Robert Solow at 9:37 P.M., December 29, 1979, in the Grand Ballroom of the Atlanta Hilton Hotel. The minutes of the meeting of August 29, 1978 were approved as published in the *American Economic Review Proceedings*, May 1979, pages 374–78.

The Secretary (C. Elton Hinshaw), Treasurer (Rendigs Fels), Managing Editor of the *American Economic Review* (George H. Borts), Managing Editor of the *Journal of Economic Literature* (Mark Perlman), and Director of *Job Openings for Economists* (Hinshaw) discussed their written reports which were available at registration and were also distributed at the meeting itself. (See their reports published in this issue.)

Prior to the reports of Borts and Perlman, Solow announced that Robert Clower had accepted the editorship of the *American Economic Review* and Moses Abramovitz that of the *Journal of Economic Literature*. The Association should congratulate itself on finding such able replacements for the current editors, who have done such outstanding jobs.

The Treasurer reported that, in view of projected deficits and continuing increases in costs as a result of inflation, the Executive Committee had voted to raise the base fee from $26.25 to $28.75 effective April 1, 1980 and submit the following recommended change in the bylaws to a mail ballot of the members: Delete the phrase "provided that the increase in any year shall not exceed ten percent" from the second paragraph of Article 1, Section 2. (See page viii of the December 1978 *American Economic Review* for bylaws.) The relevant paragraph currently reads:

> Effective January 1, 1976, the base fee is $25.00 per year. The Executive Committee may increase the dues schedule, including both the base fee and the income brackets for regular members, in proportion to the increase occurring after January 1, 1976 in relevant price and wage indexes, provided that the increase in any year shall not exceed ten percent.

If adopted it would read:

> Effective January 1, 1976, the base fee is $25.00 per year. The Executive Committee may increase the dues schedule, including both the base fee and the income brackets for regular members, in proportion to the increase occurring after January 1, 1976 in relevant price and wage indexes.

In response to a question from the audience about why the increase in publication costs exceeded predicted inflation, Fels explained that much of the difference is attributable to the one-time costs of transferring the editorial offices of both the *AER* and the *JEL* during 1980. For some period of time, four editorial offices will be in operation.

Rondo Cameron raised two objections to the fee increase: (1) Real income of members has declined, and (2) because of a perverse elasticity of demand, membership in the smaller economics associations declines when the AEA raises its dues. Fels responded that the choice faced was to raise dues or eliminate programs.

The Secretary presented the report of the Committee on Elections and the certification of the new officers for 1980 as follows:

> In accordance with the bylaws on election procedures, I hereby certify the results of the recent balloting and report the actions of the Nominating Committee, the Electoral College, and the Committee on Elections.
>
> The Nominating Committee, consisting of Franco Modigliani, Chair,

Andrew Brimmer, George G. Daly, Hirschel Kasper, Rachel McCulloch, John G. Riley, and Sherwin Rosen, submitted the nominations for Vice-Presidents and members of the Executive Committee. The Electoral College, consisting of the Nominating Committee and the Executive Committee meeting together, selected the nominee for President-elect. No petitions were received nominating additional candidates.

President-elect
William J. Baumol

Vice-Presidents	*Executive Committee*
Carl F. Christ	Jagdish N. Bhagwati
Irving B. Kravis	Martin Feldstein
H. Gregg Lewis	Robert E. Lucas, Jr.
Lionel W. McKenzie	Walter Y. Oi

The Secretary prepared biographical sketches of the candidates and distributed ballots last summer. The Committee on Elections, consisting of Ben W. Bolch, Chair, Gayle D. Riggs, and C. Elton Hinshaw, *ex officio*, canvassed the ballots and filed results with the Secretary. From the report of the Committee on Elections, I have the following information:

Number of envelopes without names for identification	135
Number of envelopes received too late	104
Number of defective ballots	2
Number of legal ballots	4,571
	4,812

On the basis of the canvass of the votes, I certify that the following persons have been duly elected to the respective offices:

President-elect (for a term of one year)
William Baumol
Vice-Presidents (for a term of one year)
Carl F. Christ
H. Gregg Lewis
Members of the Executive Committee (for a term of three years)
Martin Feldstein
Robert E. Lucas, Jr.

The Secretary presented the following resolutions, which were adopted unanimously:

BE IT RESOLVED that this meeting record a special note of appreciation to the members of the 1979 Allied Social Science Association's Convention Committee chaired by Harry Brandt. Each member has devoted many hours to planning, organizing, and conducting this ninety-second annual meeting.

BE IT RESOLVED that this meeting commend Moses Abramovitz, President-elect and 1979 Program Chair, for planning and organizing a program of wide interest and high quality.

There being no old business, the President called for new business. He stated that three resolutions by members had been submitted to the Secretary a month in advance of this meeting as required by the Association's regulations. The President called for discussion of the resolution submitted by Robert Cherry and Patrick Clawson. The resolution read:

WHEREAS the Executive Committee of the AEA has consistently forestalled debate and consideration of important social issues, such as condemnation of Nixon's decision to bomb North Vietnam (1972), condemnation of genetic theories of racial inferiority, refusal to hold AEA Convention in a non-ERA state,

WHEREAS other social science associations, without loss of their tax exempt status have adopted similar resolutions;

WHEREAS the AEA legal counsel's unresearched opinion is based on an arbitrary interpretation of the extremely vague phrase in the AEA Bylaws, "will take no partisan attitude, nor will it commit its members to any position on practical economic questions";

BE IT RESOLVED that the AEA Executive Committee's legal counsel pre-

pare a report completed by no later than April 1, 1980 which indicates what kinds, if any, of social resolutions are possible under the present Bylaws; what differences exist between our Bylaws and those of other academic associations which enable them to vote upon a broad range of resolutions and still maintain their tax exempt status; and what changes could be made, *if desired*, which would enable the AEA to vote on resolutions concerning social issues without the loss of our tax exempt status. This report will be made available upon request to any member of the Association.

Cherry spoke in favor of the resolution. He stated that its purpose is to get around the obstacle to substantive resolutions on important issues such as those cited in paragraph one of the resolution. He questioned the legitimacy of previous decisions to rule such resolutions out of order. The resolution seeks a report and information. A vote against it is a vote to supress information.

Solow noted two errors in the resolution: (1) The tax exempt status of the Association is not in question as alleged, and (2) the prohibition against a "partisan attitude" is a provision of the Articles of Incorporation, not of the bylaws. He announced that counsel has agreed to make a report on the procedure necessary to change the Articles of Incorporation and to submit a written opinion on the latitude allowed by the provision in question for the Association, as such, to take stands on social and political issues. His report will be filed with the Secretary and will be made available to members upon request. Solow further stated that whereas counsel has agreed to make such a report and whereas the "whereases" of the resolution, if adopted, would become a part of the legislative history of the resolution and provide background for interpretation of its meaning and intent, the President and the Executive Committee recommended rejection of the resolution. The resolution failed.

The President called for discussion of the resolution submitted by Heidi I. Hartmann and John V. Wells. The resolution read:

WHEREAS the Association should consider the preferences of the membership in selecting meeting and job-market sites;

WHEREAS the majority of the members present at this meeting prefer that all future meeting and job market sites be located in states which have ratified the Equal Rights Amendment (ERA) until such time as the ERA becomes an amendment to the United States Constitution:

BE IT RESOLVED that the Executive Committee abide by these preferences in selecting future meeting and job-market sites.

Hartmann spoke in favor of its adoption. She said that she was personally dismayed by the Association's meeting in Georgia, a non-ERA state. She had tried to word the resolution to overcome objections raised last year in Chicago to a similar resolution. The resolution is not political. It simply binds the Executive Committee to take into account the preferences of the members when selecting meeting sites and states what the preferences of those present at this meeting are. Henry Schloss asked if the intent of the resolution was to restrict forever the Association from meeting in states that have not ratified the Equal Rights Amendment. Hartmann responded that the restriction extended until such time as the ERA was adopted as an amendment to the U.S. Constitution. Solow noted that the announced sites for all meetings of the Association through 1985 were in states that had ratified the ERA.

After some additional discussion in which various reasons were cited for and against the resolution, Fels stated that he did not wish to vote yea nor nay. A "yea" vote would place the Association on record as taking a partisan stand, and a "nay" would imply personal opposition to ERA. He moved, and it was seconded, to postpone consideration of the resolution indefinitely.

Carolyn Shaw Bell asked that Fels withdraw his motion so that she could offer a substitute resolution. With the consent of the meeting, the motion to postpone indefinitely was withdrawn. Bell moved a substitute resolution, to wit:

> WHEREAS the Association should consider the preferences of the membership in selecting meeting and job-market sites:
>
> BE IT RESOLVED that this meeting applauds the decision of the Executive Committee to hold meetings and job-markets through 1985 in states that have ratified the ERA and affirms the wisdom of that decision.

The substitute resolution was adopted.

The President called for discussion of the resolution submitted by Ronald Calitri and Alex Azarcs. The resolution read:

> WHEREAS the Association deems the passage of the Equal Rights Amendment to be important to the status of Women in the Economics Profession;
>
> BE IT RESOLVED that the Association shall do no future business with states which have not ratified the Equal Rights Amendment, and that the Association shall cease all present business with states which have not ratified the Equal Rights Amendment as soon as alternatives can be found.

The motion to adopt the resolution failed.

The Chair then introduced Moses Abramovitz, the incoming President. There being no further business, the meeting adjourned at 11:47 P.M.

C. ELTON HINSHAW, *Secretary*

Minutes of the Executive Committee Meetings

Minutes of the Meeting of the Executive Committee in Washington, D.C., March 16, 1979.

The first meeting of the 1979 Executive Committee was called to order at 9:15 A.M. on March 16, 1979 in the Caucus Room of the Washington Hilton, Washington, D.C. The following members were present: Robert M. Solow, presiding, Henry J. Aaron, Moses Abramovitz, Irma Adelman, George H. Borts, Robert W. Clower, Rendigs Fels, Zvi Griliches, C. Elton Hinshaw, Lawrence R. Klein, Tjalling C. Koopmans, Robert J. Lampman, Marc Nerlove, Mark Perlman, and Marina v.N. Whitman. Leo J. Raskind, counsel of the Association, was also present. Present as guests for parts of the meeting were Marcus Alexis, Andrew F. Brimmer, John Chipman, Carl Christ, George G. Daly, Ann Friedlaender, Lee Hansen, Hirschel Kasper, Rachel McCulloch, Franco Modigliani, Naomi Perlman, Lloyd Reynolds, John G. Riley, Sherwin Rosen, and Wilma St. John. Solow opened the meeting by introducing the new members of the Executive Committee: Henry J. Aaron, Moses Abramovitz, Irma Adelman, and Zvi Griliches.

Minutes. The minutes of the August 28, 1978 meeting were approved as written and circulated.

Report of the Secretary (Hinshaw). The Secretary reported that the 1979 annual meetings are scheduled to be held in Atlanta, Georgia on December 28–30 with headquarters at the Atlanta Hilton Hotel. Placement services will be provided beginning December 27 and will be located in the Marriott Hotel. The schedule for subsequent meetings is: September 5–7, 1980 in Denver; December 28–30, 1981 in Washington, D.C.; and December 28–30, 1982 in New York. It was VOTED to accept the Secretary's recommendation to meet in San Francisco, December 28–30, 1983.

In an attempt to determine members' preferences concerning the timing of the annual meeting, an Informational Ballot was mailed to members in the summer of 1978. Members were asked to rank six options from one (most preferred) to six (least preferred). Three thousand, eight hundred ninety-five ballots were returned. The mean score and rank for each of the time periods was:

Time Period	Mean Score	Rank
Between Christmas & New Year's	2.74	1
Immediately after New Year's	2.99	2
Spring	3.58	3
Alternating between Christmas/ New Year's & Labor Day	3.67	4
Labor Day	3.68	5
Immediately prior to Christmas	4.01	6

Considering the results of the ballot, the low registration (about 3,000) at the Labor Day meeting in Chicago, the difficulties of organizing a separate placement meeting, and the usual difference in hotel room rates between the Labor Day and Christmas/ New Year periods, it was VOTED to terminate the experiment of alternating meeting dates and schedule future meetings during the Christmas/New Year period.

The Secretary reported that he had received sixteen petitions from graduate student associations concerning the location of the placement meetings. Thirteen read as follows:

> We, the undersigned, feel that the American Economic Association should not be holding its hiring meetings in states that have not ratified the Equal Rights Amendment.

The other three were more expansive but conveyed the same sentiment. Ann Friedlaender, Chair of the Committee on the Status of Women in the Economics Profession, spoke in favor of moving the 1979 Placement Meetings out of Atlanta. No action was taken.

Report of the Treasurer (Fels). The Treasurer reported that the Association had a surplus of $186,000 in 1978. Some of the

surplus, however, is spurious. In the transition from a hand-operated system to computerization in 1977, the mailing of second notices to members who had not paid their dues got delayed. As a result, an unusually large amount of dues applicable to 1977 (on the order of $30,000) was received too late in 1978 for inclusion in the audited financial reports for 1977 and had to be included in 1978.

The proposed budget for 1979 showed modest increases in receipts and expenditures with a surplus of about the same size as that for 1978 after the adjustment previously described. It was VOTED to approve the 1979 budget as submitted.

Report of the Editor of the American Economic Review (Borts). On recommendation of the editor, Pranab Bardhan, Peter Diamond, W. Erwin Diewert, Michael Parkin, Roy Radner, and Nancy Schwartz were elected to the Board of Editors.

Report of the Editor of the Journal of Economic Literature (Perlman). On recommendation of the editor, Robert Lampman was elected to a three-year term on the Board of Editors. It was VOTED to authorize a contract between the Association and SDC Development Corporation to market and sell on-line access to the bibliographic data base created for the *JEL* and the *Index of Economic Articles*.

Report of the Director of Job Openings for Economists (Hinshaw). The Director reported that during 1978 employers advertised 1,647 vacancies. This compares to the 1,470 new vacancies advertised in 1977. Excluding indirect cost, *JOE* continues to be self-financing.

Committee on Economic Education (Hansen). It was VOTED to support a research project on the effectiveness of the economics undergraduate major by asking the President to write a letter to chairpersons supporting the project, making the mailing list of chairpersons available to the Committee, and allocating $1,700 for postage. It was understood that additional external funding would be sought, and if not obtained, the project would be dropped.

U.S.–Soviet Exchanges (Reynolds). The fifth U.S.–Soviet Economic Symposium is scheduled to be held near New York Ci from June 10–13, 1979. The subject "Long-Term Structural Change in Nation Economics." The Soviet visit this year w presumably be followed by an invitation f a U.S. delegation to visit the Soviet Unic in 1980.

Committee on Federal Funding of Ec nomic Research (Lampman). The Committ plans to draft a full report this summer ar present a final report to the next meeting the Executive Committee in Atlanta.

Finance Committee (Fels). On behalf Robert Eisner, chairman of the Committe Fels reported that the Committee ha VOTED to increase the upper limit for t equity proposition of the Association portfolio to 75 percent. The lower limit is percent. Since the current equity share is percent, the action implied a desire to i crease the equity share.

Committee on the Status of Minorities the Economic Profession (Alexis). Alexis r ported that plans for the 1979 Summer Pr gram were virtually complete and sufficie external funding had been received to f nance the program for the next three yea There are no current financial needs. I addition, the consortium of Yale, M.I.T Michigan, Northwestern, and Stanford w be funded by the Sloan Foundation. Alexis becomes a member of the Intersta Commerce Commission, the program can b moved to one of the other consortiu schools if necessary. It was VOTED to di charge the committee appointed to evalua the program.

Honors and Awards (Adelman). Acti together as an electoral college, the Con mittee on Honors and Awards and the E ecutive Committee VOTED to award th John Bates Clark medal to Joseph Stiglitz.

1979 Program (Abramovitz). The Pres dent-elect reported that the bulk of the 197 program has been arranged. The Ely Le ture and other sessions will center on th question of the limits of welfare capitalism

Nominating Committee (Modigliani). Th Electoral College consisting of the Nomina ing and Executive Committees meetin together chose William Baumol as th nominee for President-elect, and Margar

Reid and Ronald Coase as Distinguished Fellows. The chairman of the Nominating Committee reported the following nominees for other offices: for Vice-President (two to be chosen), Carl Christ, H. Gregg Lewis, Irving Kravis, and Lionel McKenzie; for members of the Executive Committee (two to be chosen), Jagdish Bhagwati, Martin Feldstein, Robert E. Lucas, Jr., and Walter Oi.

Search Committee for Editors (Solow). It was VOTED to elect Moses Abramovitz to a three-year term (beginning January 1981) as editor of the *Journal of Economic Literature*. It was agreed that Naomi Perlman would continue as associate editor in charge of the bibliographic sections of the journal and the production of the *Index* under the general editorship of Abramovitz.

Solow sought advice about and suggestions for candidates for the editorship of the *American Economic Review*. He received same.

The meeting adjourned at 4:50 P.M.

C. ELTON HINSHAW, *Secretary*

Minutes of the Meeting of the Executive Committee in Atlanta, Georgia, December 27, 1979

The second meeting of the 1979 Executive Committee was called to order at 10:10 A.M. on December 27, 1979 in the Atlanta Hilton Hotel, Atlanta, Georgia. The following members were present: Rober Solow, presiding, Henry J. Aaron, Moses Abramovitz, Irma Adelman, George Borts, Robert Clower, Rendigs Fels, Zvi Griliches, Jack Hirshleifer, C. Elton Hinshaw, Lawrence Klein, Tjalling Koopmans, Robert Lampman, and Mark Perlman. Also present were Leo J. Raskind, Counsel, William Baumol, President-elect, and the newly elected members of the 1980 Executive Committee, Carl F. Christ, Martin Feldstein, H. Gregg Lewis, and Robert E. Lucas. Present for parts of the meeting to present reports were Marcus Alexis, Elizabeth Bailey, Donald Brown, Robert Lanzillotti, Dwight Perkins, Lloyd Reynolds, and Arnold Zellner.

Solow opened the meeting by welcoming the new members of the 1980 Executive Committee, extending the appreciation of the Association to outgoing members, and reviewing issues that had been raised with him by various members of the Association.

TABLE 1—MEMBER'S PREFERENCES FOR ANNUAL MEETING SITES
1978 Informational Ballot

City	Rank	Mean Score
San Francisco	1	1.56
Boston	2	1.92
New Orleans	3	1.94
New York	4	1.95
St. Louis	5	2.02
Denver	6	2.11
Atlanta	7	2.20
Chicago	8	2.28
Washington, D.C.	9	2.60
Philadelphia	10	2.61
Houston	11	2.78
Dallas	12	2.80
Las Vegas	13	2.89
Pittsburgh	14	3.09

Minutes. The Secretary made a correction in the minutes of March 16, 1979 which had been previously circulated. The corrected minutes were approved.

Report of the Secretary (Hinshaw). The Secretary reported that the 1980 annual meeting will be held in Denver on September 5–7 with headquarters at the Denver Hilton Hotel. Since the 1980 meeting occurs early in the academic year, the Placement Service will be held December 28–30 in Dallas, Texas. Only one Placement Service will be organized in 1980. The schedule for subsequent meetings is Washington, D.C. (1981), New York (1982), and San Francisco (1983). After reviewing possible sites for 1984 and 1985 and the results of the 1978 Informational Ballot which asked members to rate cities from one (excellent) to four (unsatisfactory) as sites for future meetings (see Table 1), the Executive Committee VOTED to approve the Secretary's request to enter into contracts with hotels in Dallas, Texas for the 1984 meeting and in New York City for 1985.

The Association will be 100 years old in 1985. Solow appointed Adelman, Borts, and Perlman to an *ad hoc* committee to consider and recommend centennial activities. The committee is to report at the March 1980

meeting of the Executive Committee. Two possibilities to mark the event were discussed: a proposal by A. W. Coats to write a history of the Association and the publication of an encyclopedia of economics. The Committee indicated a predisposition to cooperate on the Coats project but requested a detailed proposal that specifies the Association's commitment and relationship to the monograph. It was VOTED to ask Perlman to submit a proposal and budget for the encyclopedia project.

Z. A. Silberston, Secretary of the Royal Economic Society, has proposed that members of our respective organizations be offered discount subscription rates to the other one. After considering the probable effect of revenues and costs, it was VOTED not to enter into the suggested arrangement but to offer to sell bulk quantities of the AEA journals to the Royal Economic Society at average cost (FOB).

Report of the Treasurer (Fels). Nineteen seventy-nine will see the last of the series of substantial surpluses that lifted the net worth of the Association from a negative figure at the end of 1974 to three-quarters of a million dollars on September 30 of this year. The surplus expected for 1979 is well over $100 thousand. The proposed budget for 1980 shows a deficit almost as large. A dues increase is necessary, the first increase since January 1, 1978.

A large part of the deficit anticipated for 1980 is accounted for by extra expenses associated with moving the offices of the *American Economic Review* from Providence, Rhode Island, to Los Angeles, California, and part of the offices of the *Journal of Economic Literature* from Pittsburgh, Pennsylvania, to Stanford, California. For several months the Association will have to maintain two offices for each publication. Without these extra expenses, the budgeted deficit would be only 2 percent of the total expenses.

Although the main cause of the 1980 deficit will not recur, rising costs will bring about a large deficit in 1981 if dues are not increased. Since the net worth is adequate to serve as a cushion for unforeseen deficits (such as those of 1970 and 1980), there is n need for the Association to continue to hav surpluses, but the net worth is not so larg that it can afford repeated deficits. It wa VOTED to approve the 1980 budget, to rais the base membership fee to $28.75 an adjust income ranges proportionately, an to submit to the members by mail ballot a amendment to the bylaws eliminating the 1 percent constraint on dues increases in an one year.

Report of the Editor of the American Economic Review (Borts). The editor reporte that the *Papers and Proceedings* issue will g directly into page proof, bypassing gallies Much of his effort this year was focused o the problems of transferring the office to th new editor, Robert Clower, who takes ove during the fourth quarter of 1980. He calle attention to his written report (see elsewher in this issue), and thanked those members o his Board whose terms had expired.

Report of the Editor of the Journal o Economic Literature (Perlman). In additio to his written report (see elsewhere in thi issue), the editor announced that the 197 *Index of Economic Articles* was publishe and that he expects the 1976 and 197 volumes to appear in 1980. On recommendation of the editor, N. W. Balabkins Donald Harris, John Kendrick, and Isabe Sawhill were elected to the Board of Editors.

Report of the Director of Job Openings fo Economists (Hinshaw). The Director reviewed his previously circulated report (se elsewhere in this issue).

Committee on the Status of Women in th Economics Profession (Bailey). Bailey, wh chairs the Committee, noted that preliminary analysis of the AEA Universa Academic Questionnaire revealed no improvement in the status of women in th profession. In 1980 the Committee plans t allocate more funds to activities aimed specifically at increasing the number of women in tenured faculty and other responsibl positions in economics. In 1979 the Committee had spent a substantial amount o money to "professionalize" its roster o women economists and, as a consequence

had overspent its budget. It was VOTED to allocate an additional $3,500 to the Committee to cover the 1979 deficit.

Report of the Census Advisory Committee (Lanzillotti). Lanzillotti reviewed his previously circulated report (see elsewhere in this issue). It was VOTED to endorse the fellowship program instituted by the Census Bureau and to ask the Census Advisory Committee to explore with the Commissioner of Labor Statistics and the Bureau of Labor Statistics the possibility and desirability of establishing an AEA advisory committee to the Bureau.

Committee of U.S.–Soviet Exchanges (Reynolds). Reynolds reported that another exchange visit to the Soviet Union had been arranged for 1980, and that he expects the visit to be funded by the U.S. International Communications Agency. He also reported on the 1979 exchange visit to the United States.

Committee on the Status of Minority Groups (Alexis). Alexis, who chairs the Committee and administers the summer program, reviewed briefly the sixth summer program. Upon his ascension to Commissioner of the Interstate Commerce Commission, it will be necessary to find another Director and a new home for the program. Yale University is willing to host the program and has made arrangements to do so provided the Association will be responsible for the direct costs. He projected a total cost over three years of $297,000. The Sloan Foundation and others have indicated interest. He wished to defer consideration of the budget until the March meeting of the Executive Committee. It was VOTED to support the move of the program to Yale University for a period of at least three years and exhort and help the Committee on the Status of Minorities in the Economics Profession to make its best effort in seeking funds for the program.

Committee on Federal Funding (Lampman). He reviewed his written report (see elsewhere in this issue). It was VOTED to offer to appoint committees to advise federal departments and agencies with respect to their primary data collection and presentation of basic economic indicators. It was understood that caution would be taken in appointing such committees and that only one or two would be appointed in the first instance. Additional committees might be appointed if the experience proved useful. It was VOTED to appoint a standing committee to serve as a channel of communication with the Economics Program of the Social Science Division of the National Science Foundation.

Finance Committee (Fels). On behalf of the Committee, Fels reported that Stein, Roe, and Farnham, the Association's investment counselor, had raised their fee to 3/8 of 1 percent. He indicated that, even with the increase, their fee was below that of competitors and recommended continuation with them. There was no objection.

1980 *Program* (Baumol). Baumol, President-elect and Program Chair for the 1980 meetings, stated that the program outline was pretty much set except for individual presenters and presiders. To the extent that there is a theme, it is "Applied Theory."

National Bureau of Economic Research (Christ). Christ, the Association's representative on the Board of the NBER, asked for instructions. He was advised to use his own judgment.

Resolutions of Members to the Annual Business Meeting (Solow). The resolutions submitted by members for consideration at the annual business meeting were reviewed and discussed.

The meeting was adjourned at 5:50 P.M.

C. ELTON HINSHAW, *Secretary*

Report of the Secretary for 1979

Annual Meetings. In 1980 the annual meetings will be held at the Denver Hilton Hotel in Denver, Colorado on September 5–7. The schedule for subsequent meetings is December 28–30, 1981 in Washington, D.C., December 28–30, 1982 in New York, December 28–30, 1983 in San Francisco, December 28–30, 1984 in Dallas, and December 28–30, 1985 in New York.

Employment Services. For those meetings scheduled for December 28–30, employment services will be provided at the annual meeting but will begin December 27. Because the 1980 annual meetings will occur at an early period in the academic year, it was decided to provide employment services at a later time. The 1980 placement meeting will be held in Dallas, Texas during December 28–30. Placement services will not be organized for the September meetings in Denver.

The National Registry for Economists continues to be operated on a year-round basis by the Illinois State Employment Service. Economists looking for jobs and employers are urged to register. This is a placement service that maintains the anonymity of employers. The Association is indebted to the Registry for assistance and supervision of the employment service provided at the annual meetings.

Employers are reminded of the Association's bimonthly publication, *Job Openings for Economists*, and their professional obligation to list their openings.

Membership. The total number of members and subscribers, shown in Table 1, reached an all-time high of 26,787 at the end of 1975. After declining for two years, the total increased in 1978 and again in 1979.

Table 1—Members and Subscribers
(End of Year)

	1977	1978	1979
Class of Membership			
Annual	14,379	15,698	16,203
Junior	1,731	1,857	1,884
Life	375	389	388
Honorary	33	35	35
Family	284	307	315
Complementary	537	615	634
Total Members	17,339	18,901	19,459
Subscribers	6,728	6,893	6,963
Total Members and Subscribers	24,067	25,794	26,422

Permission to Reprint and Translate. Official permissions to quote from, reprint, or translate and reprint articles from the *American Economic Review* and the *Journal of Economic Literature* totaled 262 in 1979 compared to 303 in 1978. Upon receipt of a request for permission to reprint an article, the publisher or editor making the request is instructed to get the author's permission in writing and send a copy to the Secretary as a condition for official permission. The Association suggests that authors charge a fee of $150, but they may charge some other amount, enter into a royalty arrangement, waive the fee, or refuse permission altogether.

Committees and Representatives. Listed below are those who served the Association during 1979 as members of Committees or representatives. The year in parenthesis indicates the final year of the term to which they have been appointed most recently. On behalf of the Association, I wish to thank them all for their services.

Ad Hoc Advisory Committee to the National Commission on Employment and Unemployment Statistics
Harold Watts, *Chair*
Orley Ashenfelter
Carolyn Shaw Bell
Charles C. Holt

Ad Hoc Committee on Publishing Contracts
Martin Shubik, *Chair*
William J. Baumol
Leo Raskind
C. Elton Hinshaw, *ex officio*

AD HOC COMMITTEE TO REVIEW NEW PROPOSED STANDARD OCCUPATIONAL CLASSIFICATION SYSTEM
H. Gregg Lewis, *Chair*
Victor Fuchs
Margaret S. Gordon
Michael Piore
Sherwin Rosen

AD HOC COMMITTEE TO STUDY SGE REPORT
Henry J. Aaron, *Chair*
George Jaszi
Hyman Kaitz

BUDGET COMMITTEE
Henry J. Aaron (1981)
Marina v.N. Whitman (1980)
Robert J. Lampman (1979), *Chair*
Rendigs Fels, *ex officio*
Moses Abramovitz, *ex officio*
Robert Solow, *ex officio*

CENSUS ADVISORY COMMITTEE
Ann D. Witte (1979)
Otto Eckstein (1980)
Victor R. Fuchs (1980)
George L. Perry (1980)
Norman Simler (1980)
Lester C. Thurow (1980)
Carolyn Shaw Bell (1979)
Andrew F. Brimmer (1979)
Burton Malkiel (1979)
Arnold Zellner (1979), *Chair*
Barbara Bergmann (1981)
Martin H. David (1981)
Robert F. Lanzillotti (1981)
William Niskanen (1981)
Richard D. Karfunkle (1981)

COMMITTEE ON ECONOMIC EDUCATION
Allen C. Kelley (1979), *Chair*
George Leland Bach (1980)
William E. Becker (1980)
Karl E. Case (1981)
W. Lee Hansen (1981)
Walter Heller (1979)
Keith Lumsden (1980)
John Siegfried (1981)
Rendigs Fels, *ex officio*

ECONOMICS INSTITUTE POLICY AND ADVISORY BOARD
Edwin S. Mills (1981), *Chair*
Douglass C. North (1982)
Dwight Perkins (1982)
G. Edward Schuh (1982)
Axel Leijonhufvud (1981)
Carlos F. Diaz Alejandro (1980)
Raymond Vernon (1980)
Carl Keith Eicher (1979)
Anne O. Krueger (1979)

COMMITTEE ON ELECTIONS
Ben Bolch (1980), *Chair*
Gayle D. Riggs (1981)
C. Elton Hinshaw, *ex officio*

COMMITTEE ON FEDERAL FUNDING OF ECONOMIC RESEARCH
Robert J. Lampman, *Chair*
Richard Freeman
Anne O. Krueger
George L. Perry
Richard N. Rosett

FINANCE COMMITTEE
Robert Eisner (1980), *Chair*
Robert G. Dederick (1979)
James Lorie (1981)
Rendigs Fels, *ex officio*

COMMITTEE ON HONORARY MEMBERS
Leonid Hurwicz (1980), *Chair*
Hendrik S. Houthakker (1984)
George J. Stigler (1984)
William J. Baumol (1982)
Hollis B. Chenery (1982)
Paul A. Samuelson (1980)

COMMITTEE ON HONORS AND AWARDS
John Chipman (1980), *Chair*
Daniel McFadden (1984)
Oliver E. Williamson (1984)
Moses Abramovitz (1982)
Carl F. Christ (1982)
Dale T. Mortensen (1982)
James W. McKie (1980)

NOMINATING COMMITTEE (1979)
Franco Modigliani, *Chair*
Andrew F. Brimmer

George Daly
Hirschel Kasper
Rachel McCulloch
John Riley
Sherwin Rosen

COMMITTEE ON POLITICAL DISCRIMINATION
Carl M. Stevens (1980), *Chair*
Anne P. Carter (1981)
Kenneth J. Arrow (1980)
John G. Gurley (1980)
Robert E. Lucas, Jr. (1980)
Harold Barnett (1981)

COMMITTEE ON PUBLICATIONS
Robert Ferber (1981), *Chair*
Martin Bronfenbrenner (1981)
Kenneth W. Leeson (1981)
Barbara Reagan (1981)
Edwin Burmeister (1979)
Peter A. Diamond (1979)
C. Elton Hinshaw, *ex officio*

SEARCH COMMITTEE FOR EDITORS
Robert Solow, *Chair*
Irma Adelman
Albert Ando
George R. Feiwel
Martin S. Feldstein
John G. Gurley
Michael C. Lovell
Bernard Saffran

COMMITTEE ON THE STATUS OF MINORITY GROUPS IN THE ECONOMICS PROFESSION
Marcus Alexis (1980), *Chair*
James N. Morgan (1981)
Guy H. Orcutt (1980)
Albert Rees (1980)
Andrew F. Brimmer (1979)
Alice Rivlin (1979)

COMMITTEE ON THE STATUS OF WOMEN IN THE ECONOMICS PROFESSION
Ann Friedlaender (1979), *Chair*
Barbara A. Jones (1981)
Helen F. Ladd (1981)
Marianne Ferber (1980)
Ruth Gilbert Shaeffer (1980)
Miriam K. Chamberlin (1979), *Vice-Chair*
William F. Hellmuth (1979)
Robert Solow, *ex officio*

COMMITTEE ON U.S.–CHINA EXCHANGES
Dwight H. Perkins (1981), *Chair*
Hollis Chenery (1981)
Gregory Chow (1981)
Robert Dernberger (1981)
John G. Gurley (1981)
Lawrence R. Klein (1981)
Tjalling Koopmans (1981)
Robert Solow (1981)
Benjamin Ward (1981)

COMMITTEE ON U.S.–SOVIET EXCHANGES
Lloyd G. Reynolds (1979), *Chair*
Abram Bergson (1979)
John Meyer (1979)
Rendigs Fels, *ex officio*

Council and Other Representatives

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE SECTION K ON SOCIAL AND ECONOMIC SCIENCES
William Nordhaus (1979)

AMERICAN COUNCIL OF LEARNED SOCIETIES
C. Elton Hinshaw (1981)

FEDERAL STATISTICS USERS CONFERENCE
Paul Wonnacott (1979)

INTERNATIONAL ECONOMIC ASSOCIATION
Anne Krueger (1984)
Franco Modigliani (1981)

INTERSOCIETY COMMITTEE ON TRANSPORTATION
William Dodge

POLICY BOARD OF THE JOURNAL OF CONSUMER RESEARCH
Lester Telser (1982)

NATIONAL ARCHIVES ADVISORY COUNCIL–GENERAL SERVICES ADMINISTRATION
William N. Parker (1981)

NATIONAL BUREAU OF ECONOMIC RESEARCH
Carl F. Christ (1981)

SOCIAL SCIENCE RESEARCH COUNCIL
Finis Welch (1981)

SSRC–COMMITTEE OF PROFESSIONAL ASSOCIATIONS ON FEDERAL STATISTICS (COPAFS)
John H. Cumberland (1980)
Gary Fromm (1979)

Representatives of the Association on Various Occasions–1979

INAUGURATIONS
Earl Robert Mezoff, Tusculum College
Herman A. Stribling, Jr.

Bill D. Draper, Point Loma College
Roger Frantz

Harry Edmund Smith, Austin College
Edmund Lacy

Charles Waymond Merideth, Atlanta University Center, Inc.
Tate Whitman

Jushua L. Smith, Manhattan Community College of CUNY
M. Ishaq Nadiri

Jerry M. Anderson, Ball State University
Bernerd Bogar

Richard Allan Terry, Quinnipiac College
James C. Hsiao

Sister Emmanuel Renner, College of Saint Benedict
Joseph M. Friedrich

Eugene Sumner Mills, Whittier College
Paul T. Mu

1979 ASSA Convention Committee

Harry Brandt, *Chair*
John M. Godfrey, *Vice-Chair*
Jerry Adams
James C. Armstrong
Janet Aschenbrenner
William E. Black
Kong Chu
James E. Clark
Caroline P. Clotfelter
Donald M. Cruse
Arnold A. Dill
Paul G. Farnham
M. M. Galloway
Marilu McCarty
W. Bethel Minter
Janet S. Rankin
Cindy Russell
Charles D. Salley
Arthur F. Schreiber
Jack R. Sicard
Violet O. Sikes
Barbara Weaver
Philip M. Webster
Melissa Williams
Mary L. Winer

1980 ASSA Convention Committee

Thomas E. Davis, *Chair*
Glenn Miller, *Vice-Chair*
Tucker Hart Adams
Norma J. Ayres
Kathleen M. Cooper
Doris M. Drury
Susan F. Norman
Barry K. Robinson
Violet O. Sikes
Barbara Weaver
Melissa Williams
Mary L. Winer

C. Elton Hinshaw, *Secretary*

Report of the Treasurer for the Year Ending December 31, 1979

Nineteen seventy-nine saw the last of the series of substantial surpluses that lifted the net worth of the American Economic Association from a negative figure at the end of 1974 to about three-quarters of a million dollars now. At this writing, audited figures for 1979 are not yet available. They can be expected to differ from the expected results shown in Table 1. Nevertheless, the expected surplus for 1979 of over $100 thousand is accurate enough for purposes of the present discussion.

The budget shown in Table 1 for 1980 is substantially the same as the one approved by the Executive Committee on December 27, 1979. (It incorporates minor changes which will be submitted to the Executive Committee at its meeting on March 21, 1980, and other small changes may be expected later.) The expected deficit for 1980, though smaller than the expected surplus for 1979, is substantial, but is entirely accounted for by extra expenses associated with moving the offices of the *American Economic Review* from Providence, Rhode Island, to Los Angeles, California, and moving part of the offices of the *Journal of Economic Literature* from Pittsburgh, Pennsylvania, to Stanford, California. For several months, the Association will have to maintain duplicate offices for each publication. Without these extra expenses, the budgeted deficit would be small.

The net worth, even after 1980, will be adequate to serve as a cushion for unforeseen developments, but it is not so large that the Association can afford repeated deficits. Although the cause of the 1980 deficit will not recur, rising costs would bring about a large deficit in 1981 if dues are not increased. Ideally, the dues rate for regular members would be raised from $26.25 (the base rate which went into effect on January 1, 1978) to $30 on January 1, 1981, with dues for other membership categories and subscription rates raised in the same proportion. This would be an increase of 14 percent over a three-year period, a compound rate of a little over 4 percent per year. The present bylaws, however, permit the Executive Committee to raise dues no more than 10 percent in any one year. To make possible an increase in the base rate to $30 effective January 1, 1981, the Executive Committee at its meeting on December 27, 1979, voted to submit a change in the bylaws to the membership for approval or otherwise by mail ballot at the same time as the election ballots in the fall of 1980. Under the proposed change the 10 percent per year limit would be abolished in favor of a limit based solely on the increase in relevant price and wage indexes. The Executive Committee would not be allowed to increase dues more than in proportion to the general rate of inflation.

Since the outcome of the balloting could not be known, the Budget Committee recommended, and the Executive Committee approved, an increase in the base rate to $28.75 effective April 1, 1980. This increase is slightly less than 10 percent over the rate in effect since January 1, 1978 (and less than 5 percent per year). This action has the effect of raising the dues for other membership categories in the same proportion. It will have only a limited effect on 1980 revenues, since about half the members and nearly all subscribers are on a calendar year basis. The effect of this increase is included in Table 1. Without it the budgeted deficit would be $20 thousand more. If the proposed change in the bylaws is approved, I expect to recommend that the base rate be further raised to $30 effective January 1, 1981. If it is disapproved, I shall recommend raising it to $31 effective April 1, 1981. Whether either action will be enough to head off a deficit in 1981 cannot be foretold, but in any event dues will have to be

TABLE 1—AMERICAN ECONOMIC ASSOCIATION, BUDGET, ACCOUNTING BASIS, 1980
(Thousands of dollars)

	1978 (Actual)	1979 Budget (3-16-79)	1979 Expected (12-31-79)	1980 Budget (12-27-79)
REVENUE				
Operating Income				
Dues and subscriptions	903[a]	880	840	870
Advertising	87	90	93	90
JOE subscriptions	16	20	20	20
Sales—Miscellaneous	33	65	45	35
Sales—Mailing list	41	40	40	40
Sales—*Index*	49	60	77	50
Annual meeting	1	0	3	6
Other income	20	20	23	25
Total Operating Income	1,150	1,175	1,141	1,136
Investment Income				
Interest and dividends	61	57	73	80
Real capital gains (losses)	(22)	(22)	(40)	(24)
Total Investment Income	38	35	33	56
TOTAL REVENUE	1,189	1,210	1,174	1,192
EXPENSES				
Publications				
American Economic Review	303	307	310	355
Journal of Economic Literature	338	353	372	420
Directory	50	55	55	60
Job Openings for Economists	28	30	32	35
Index of Economic Articles	20	20	35	23
Start-up of new publication offices	0	0	0	77
Total Publications Expense	739	765	804	970
Operating and Administrative				
Salaries	107	114	112	129
Rent	10	10	10	11
Mailing list	38	38	28	31
Accounting and legal	9	11	10	11
Office supplies	11	12	8	9
Postage	15	18	16	18
Miscellaneous	30	27	27	29
Committees				
Women	9	9	12	11
Minorities	11	9	9	11
Political discrimination	1	9	0	0
Other	10	12	11	12
Annual meeting	4	6	6	7
Federal income tax	12	13	13	13
Total Operating and Administration	267	288	262	292
TOTAL EXPENSES	1,006	1,053	1,066	1,262
SURPLUS (DEFICIT)	182	157	108	(70)

[a]The 1978 figures for dues and subscriptions is overstated by an unknown amount. In the transition from a hand-operated system to computerization in 1977, the mailing of second notices to members who had not paid their dues was delayed. As a result, an unusually large amount of dues applicable to 1977 was received too late in 1978 for inclusion in the audited financial reports for the previous year and had to be included in the 1978 data. The average amount of dues and subscription receipts for 1977 and 1978 was $823 thousand. The correct amounts were probably somewhat smaller than that in 1977 and somewhat larger in 1978.

increased annually from now on unless inflation subsides or the Association decides to curtail its activities.

At the annual meeting on December 29, 1979, several members expressed strong opposition to dues increases. In particular, Rondo Cameron pointed out that real incomes of economics professors have been declining. He also said that dues increases by the American Association have an adverse effect on membership and revenues of other economics societies such as the Economic History Association.

Since the two managing editors and the Secretary have operated with a high degree of efficiency, the alternative to raising dues would be to curtail the activities of the Association. Limited one-time economies could be achieved by reducing the size of the principal journals, and a large reduction in expenses could be effected by abolishing one of them. It is hard to believe that reducing activities would serve the best interest of the profession.

When the proposed change in the bylaws is submitted to the members for approval, the ballots will be accompanied by a short explanation of the Executive Committee's reasons for proposing it. If opponents of the change so desire, a short statement of the arguments against it can be included also.

RENDIGS FELS, *Treasurer*

Report of the Finance Committee*

The accompanying inventory summary lists the securities held by the American Economic Association as of December 31, 1979, with costs and market values as of that date. The total market value of the securities portfolio at year-end was $1,536,847. After making adjustments for cash additions and withdrawals, we estimate that the Association's investment portfolio experienced a total investment return of +17.9 percent during 1979.

In past years, the portfolio included the funds remaining from a Special Grant made by the Ford Foundation in January of 1969 and subsequently commingled with the Association's account. However, during 1979, the Special Grant portion of the portfolio was closed out. Therefore, the entire $1,536,847 represented Association funds as of December 31, 1979.

At its annual meeting in late 1978, the Finance Committee approved an investment policy of establishing an equity ratio range of 50–75 percent. Moreover, the investment manager was authorized to lengthen fixed-income maturities at his discretion.

Mindful of this overall policy, some portfolio restructuring was undertaken last year, including the purchase of new commitments in Atlantic Richfield and Northern Telecommunications and the addition to a number of existing portfolio holdings. One position (Proctor & Gamble) was eliminated. As a result, the portfolio's overall equity exposure was increased during 1979. With respect to fixed-income securities, the greatest emphasis was placed on one- to five-year maturities as some maturity lengthening was undertaken.

In terms of the portfolio's investment experience, the Committee can also report that, in addition to the full portfolio's return of +17.9 percent for the year, the Association's equities taken alone had a total return of +23.9 percent which was substantially greater than that of the widely followed market averages (total return for the Dow Jones Industrial Average and the *S & P 500* was +10.4 percent and +18.5 percent, respectively). In terms of the portfolio's fixed-income securities, it is worth noting that this component's total return was +9.5 percent which compares quite favorably with the declines registered by the major bond market averages which are strongly influenced by longer-term maturities.

Looking into 1980, we took cognizance of projections of a moderate consumer-led recession of six to nine months duration—followed by a historically weak recovery in 1981—with inflation remaining high but gradually decelerating, with a cyclical peaking in interest rates, and with the dollar stabilizing and becoming firmer. While it is our collective judgment that this economic pattern represents the most likely course of events, it should be noted that all of the clouds which menaced and ultimately engulfed a similar outlook at the beginning of 1979 are still present.

With respect to the prospects for common stocks, it may be helpful to examine the surprising 1979 stock market experience which found equities rising broadly despite a confluence of unfavorable developments. It is our judgment that the most logical explanation for this market strength in the face of an abundance of "bad news" is that investors are now increasingly beginning to recognize that under certain conditions equities may indeed be a good hedge against inflation. Over time, common stocks will prove to be an effective hedge against inflation when corporate earnings and dividends are rising more rapidly than prices. The evidence shows that, from the end of 1974 to the end of 1979, inflation compounded at 7–8 percent annually. However, what has received less publicity is the fact that corporate profitability has grown at an even faster

*The Report of the Finance Committee is informational and is not an audited financial statement. Consequently, there may be some discrepancies between figures in the Report of the Finance Committee and the Auditors' Report which follows.

TABLE 1—INVENTORY SUMMARY AS OF DECEMBER 31, 1979

	Value	Percent	Estimated Income	Estimated Current Yield
Cash Equivalents	$312,136	20.3	$39,955	12.8
Short-Term Securities	184,935	12.0	18,425	10.0
Medium-Term Securities	75,875	4.9	8,063	10.6
Long-Term Securities and Preferred Stocks	24,621	1.6	2,813	11.4
Convertible Securities	0	0.0	0	0.0
Equity Securities	939,270	61.1	36,422	3.9
TOTAL	1,536,847	100.0	105,678	6.9

TABLE 2—INVENTORY AND APPRAISAL AS OF DECEMBER 31, 1979

	Amount	Price	Value	Unit Cost	Total Cost	Estimated Income
Cash Equivalents and Fixed-Income Securities (38.9 percent)						
CASH EQUIVALENTS (0–1 YEAR) (52.2 percent)						
Cash			$7		$7	$1
Stein Roe Cash Reserves, Inc.	312,138	1	312,139	1	312,140	39,954
Subtotal Cash Equivalents			312,146		312,147	39,955
Other Short-Term Securities (1–5 Years) (31.0 percent)						
Fed. Nat. Mtg Assn (10.000 04/10/81)	50,000	98	49,110	99	49,281	5,000
U.S. Treasury Notes (10.125 09/30/81)	40,000	98	39,325	100	39,940	4,050
Fed. Nat. Manuf. Assn (9.500 03/10/83)	50,000	96	48,125	98	49,141	4,750
U.S. Treasury Notes (9.25 05/15/84)	50,000	97	48,375	96	48,094	4,625
	190,000		184,935		186,456	18,425
Subtotal Other Short-Term Securities			184,935		186,456	18,425
Medium-Term Securities (5–10 Years) (12.7 percent)						
Fed. Farm Cr. Banks (10.750 10/20/86)	50,000	100	50,219	100	49,859	5,375
U.S. Treasury Notes (10.750 11/15/89)	25,000	103	25,656	100	24,955	2,688
	75,000		75,875		74,815	8,063
Subtotal Medium-Term Securities			75,875		74,815	8,063
Long-Term Securities (More than 10 years) (4.1 percent)						
Hydro-Quebec Debentures (11.250 10/15/09)	25,000	98	24,621	100	24,906	2,813
Subtotal Long-Term Securities	25,000		24,621		24,906	2,813
Total Cash and Fixed-Income Securities			597,577		598,323	69,256
Equity Securities (61.1 percent)						
Utilities (3.0 percent)						
Central and Southwest	2,000	14	28,250	12	24,556	2,840
Banks (7.5 percent)						
Continental Illinois	1,200	29	34,800	30	35,781	1,920
First Bank System	900	40	35,550	27	24,070	1,836
			70,350		59,851	3,756
Other Financial (3.5 percent)						
Alexander and Alexander	1,000	33	32,750	9	9,325	1,400
Foods and Related (3.1 percent)						
Philip Morris	800	36	28,800	22	17,726	1,000
Paper and Forest Products (3.9 percent)						
Fort Howard Paper	800	45	36,301	41	32,674	1,056
Machinery and Construction (4.2 percent)						
Deere	1,000	39	39,126	33	32,554	1,800
Energy (25.2 percent)						
Atlantic Richfield	1,000	80	80,000	71	70,536	2,800

TABLE 2—CONTINUED

	Amount	Price	Value	Unit Cost	Total Cost	Estimated Income
Cities Service	600	84	50,100	51	30,773	2,160
Conoco Inc.	800	47	37,800	19	15,580	1,520
Gulf Oil	800	35	27,700	17	13,321	1,800
Mapco	500	37	18,313	18	8,855	700
Standard Oil Ohio	255	88	22,504	40	10,076	,408
			236,417		149,141	9,388
Oil Service (13.3 percent)						
Halliburton	500	85	42,500	64	31,897	1,000
Ocean Drilling and Exploration	950	86	81,938	42	39,699	1,425
			124,438		71,596	2,425
Drugs and Medical (8.2 percent)						
Abbott Lab	1,000	41	41,125	21	21,360	1,000
Merck	500	72	36,125	57	28,402	1,150
			77,250		49,762	2,150
Electrical Products (7.0 percent)						
General Electric	690	51	34,932	36	24,536	1,932
Northern Telecom Ltd	700	43	30,363	44	30,569	,700
			65,295		55,105	2,632
Computers and Office Equipment (3.3 percent)						
IBM	480	64	30,901	28	13,325	1,651
Broadcasting and Publishing (2.8 percent)						
CBS	500	52	26,188	37	18,662	1,400
Miscellaneous (15.3 percent)						
Corning Glass	400	56	22,400	60	23,984	848
Disney	700	45	31,413	22	15,503	504
Eastman Kodak	600	48	28,876	65	38,911	1,740
McDonalds	700	43	30,364	48	33,402	392
Minnesota Mining and Mfg	600	50	30,151	54	32,473	1,440
			143,204		144,273	4,924
TOTAL EQUITY SECURITIES			939,270		678,550	36,422
TOTAL SECURITIES CASH			1,536,847		1,276,873	105,678

pace over this same time horizon. For example, nonfinancial corporate after-tax cash flow, perhaps the best gauge of true overall corporate profitability and the best measure of corporate dividend-paying ability, increased almost twice as rapidly as inflation, recording a compound annual growth rate of 15.2 percent. With all of the negative surprises dominating the headlines in 1979, not much attention has been paid to the fact that after-tax corporate profits rose approximately 21 percent last year. When viewed from this perspective, the impressive experience of 1979, with equities rising broadly in the face of the relentless succession of disquieting developments, become more understandable and indeed more logical.

In terms of investment policy, balancin our long-term enthusiasm against the realit of present problems, the Committee decide to maintain its guideline of keeping the pro portion of the portfolio in equities betwee 50 and 75 percent. Moreover, the Com mittee agreed that the average maturity o interest-bearing securities other than a re serve for liquidity needs should be in creased, but specified that it should not ex ceed eight years.

ROBERT EISNER, *Chai*

Auditors' Report

To the Executive Committee of
The American Economic Association:

We have examined the statement of assets and liabilities of THE AMERICAN ECONOMIC ASSOCIATION (a District of Columbia corporation, not for profit) as of December 31, 1979 and 1978, and the related statements of revenues and expenses, changes in general and restricted fund balances and changes in assets and liabilities for the years then ended. Our examinations were made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

In our opinion, the accompanying financial statements present fairly the assets and liabilities of The American Economic Association as of December 31, 1979 and 1978, and its revenues and expenses, changes in fund balances and the changes in its assets and liabilities for the years then ended, in conformity with generally accepted accounting principles consistently applied during the periods.

Arthur Andersen & Co.

Nashville, Tennessee
February 20, 1980

The American Economic Association Statement of Assets and Liabilities
December 31, 1979 and 1978

Assets	1979	1978
Cash	$ 121,183	$ 87,860
Investments, at market (Notes 1 and 2):		
Temporary investments	–	431,092
Permanent investments	1,541,376	902,093
	1,541,376	1,333,185
Accounts Receivable:		
Advertising, back issues, etc.	108,859	111,886
Allowance for doubtful accounts	(404)	(2,318)
	108,455	109,568
Inventory of *Index of Economic Articles*, at cost	37,268	52,837
Prepaid Expenses	17,427	17,056
Office Furniture and Equipment, at cost, less accumulated depreciation of $10,375 in 1979 and $8,508 in 1978	13,298	14,674
Total Assets	$1,839,007	$1,615,180

Liabilities and Fund Balances	1979	1978
Accounts Payable and Accrued Liabilities	$ 217,408	$ 313,060
Deferred Income (Note 1):		
Life membership dues	60,168	62,790
Other membership dues	357,446	330,986
Subscriptions	190,084	175,973
Job Openings for Economists	13,634	12,756
	621,332	582,505
Accrual for Directory (Note 1)	80,689	30,145
Fund Balances:		
Restricted (Note 4)	6,011	58,774
Add (deduct)—Unrecognized change in market value of investments (Notes 3 and 4)	–	(2,584)
	6,011	56,190
General	876,786	674,494
Add (deduct)—Unrecognized change in market value of investments (Notes 1 and 3)	36,781	(41,214)
General fund-net worth	913,567	633,280
Total fund balances	882,797	733,268
Add (deduct)—Unrecognized change in market value of investments (Notes 1 and 3)	36,781	(43,798)
Net fund balance	919,578	689,470
Total Liabilities and Fund Balances	$1,839,007	$1,615,180

The accompanying notes to financial statements are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF REVENUES AND EXPENSES FOR THE YEARS ENDED DECEMBER 31, 1979 AND 1978

	1979	1978
REVENUES FROM DUES AND ACTIVITIES:		
Membership dues and subscriptions	$ 554,502	$ 612,504
Nonmember subscriptions	287,285	289,947
Job Openings for Economists subscriptions	19,281	16,021
Advertising	97,939	87,350
Sale of *Index of Economic Articles*	77,299	48,980
Sale of copies, republications, and handbooks	45,277	33,411
Sale of mailing list	34,246	41,240
Annual meeting	3,523	727
Sundry	21,673	20,099
	1,141,025	1,150,279
INVESTMENT GAINS (Note 2)	71,388	38,303
Net revenues	**1,212,413**	**1,188,582**
PUBLICATION EXPENSES:		
American Economic Review	314,635	302,640
Journal of Economic Literature	376,797	338,284
Directory publication (Note 1)	55,000	50,000
Job Openings for Economists	33,616	28,413
Index of Economic Articles	40,716	19,983
	820,764	739,320
OPERATING AND ADMINISTRATIVE EXPENSES:		
General and administrative—		
Salaries	107,635	106,738
Rent	10,137	10,086
Other (Exhibit I)	83,848	103,147
Committee	31,260	30,830
Annual meeting	3,889	3,830
Provision for federal income taxes (Note 6)	14,300	12,400
	251,069	267,031
Total expenses	**1,071,833**	**1,006,351**
REVENUES IN EXCESS OF EXPENSES	**$ 140,580**	**$ 182,231**

The accompanying notes to financial statements and Exhibit I are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF CHANGES IN GENERAL FUND BALANCE FOR THE YEARS ENDED DECEMBER 31, 1979 AND 1978

	Total	Operations	Market Value Adjustments
Balance at January 1, 1978	**$451,602**	**$184,157**	**$267,445**
Add—market value adjustments resulting from inflation (Note 1)	40,661	–	40,661
Add—revenues in excess of expenses	182,231	182,231	–
Balance at December 31, 1978	**$674,494**	**$366,388**	**$308,106**
Add—market value adjustments resulting from inflation (Note 1)	61,712	–	61,712
Add—revenues in excess of expenses	140,580	140,580	–
Balance at December 31, 1979	**$876,786**	**$506,968**	**$369,818**

The accompanying notes to financial statements are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF CHANGES IN RESTRICTED FUND BALANCES FOR THE YEAR ENDED DECEMBER 31, 1979

	Balance at January 1	Receipts	Disbursements	Allocation of Investment Gains (Note 4)	Balance at December 31
The Ford Foundation grant for Economics Institute's orientation program for foreign graduate students of economics	$50,721	$ –	$(60,195)	$9,474	$ –
The Alfred P. Sloan Foundation, Chase Manhattan Bank, and Ford Foundation grants for increase of educational opportunities for minority students in economics	–	67,335	(67,335)	–	–
Funds reserved by the Association for publication of revised editions of *Graduate Study in Economics,* a guide originally published with funds from a Ford Foundation grant	–	2,770	(2,770)	–	–
The Asia Foundation grant for Asian economists' membership dues to the American Economic Association and related travel expenses	466	–	(466)	–	–
The Minority scholarship fund for minority students applying for graduate work in economics	5,000	–	–	–	5,000
The Ford Foundation grant for development of a consortium on graduate studies in economics for minorities	1,676	–	(1,676)	–	–
The International Communication Agency grant arranging for a delegation of American economists to participate in a joint symposium with the Soviet Federation of Economic Institutions	–	7,130	(7,130)	–	–
Sundry	911	100	–	–	1,011
	$58,774	$77,335	$(139,572)	$9,474	$6,011

The accompanying notes to financial statements are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF CHANGES IN RESTRICTED FUND BALANCES FOR THE YEAR ENDED DECEMBER 31, 1978

	Balance at January 1	Receipts	Disbursements	Allocation of Investment Gains (Note 4)	Balance at December 31
The Ford Foundation grant for Economics Institute's orientation program for foreign graduate students of economics	$47,305	$ –	$ –	$3,416	$50,721
The Alfred P. Sloan Foundation, Chase Manhattan Bank and Ford Foundation grants for increase of educational opportunities for minority students in economics	599	57,810	(58,409)	–	–
Funds reserved by the Association for publication of revised editions of *Graduate Study in Economics,* a guide originally published with funds from a Ford Foundation grant	–	5,915	(5,915)	–	–
The Asia Foundation grant for Asian economists' membership dues to The American Economic Association and related travel expenses	616	–	(150)	–	466
The Minority scholarship fund for minority students applying for graduate work in economics	5,000	–	–	–	5,000
The Ford Foundation grant for development of a consortium on graduate studies in economics for minorities	(1,220)	10,000	(7,104)	–	1,676
The International Communication Agency grant arranging for a delegation of American economists to participate in a joint symposium with the Soviet Federation of Economic Institutions	–	14,219	(14,219)	–	–
Sundry	811	100	–	–	911
	$53,111	$88,044	$(85,797)	$3,416	$58,774

The accompanying notes to financial statements are an integral part of this statement.

THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF CHANGES IN ASSETS AND LIABILITIES FOR THE YEARS ENDED DECEMBER 31, 1979 AND 1978

	1979	1978
Cash, beginning of year	**$ 87,860**	**$121,831**
SOURCE (USE) OF FUNDS:		
Revenues in excess of expenses	140,580	182,231
Add—noncash charges—		
Depreciation	1,295	1,231
Directory publication (Note 1)	55,000	50,000
Market value adjustments (Note 1)	11,430	22,397
Funds provided by operations	208,305	255,859
(Increase) decrease in—		
Receivables and prepaid expenses	742	(21,146)
Inventory of *Index of Economic Articles*	15,569	(21,765)
Investments	(208,191)	(328,185)
Office furniture and equipment	80	(8,168)
Increase (decrease) in—		
Accounts payable and accrued liabilities	(100,108)	(85,080)
Deferred income	38,827	171,563
Restricted funds	(52,763)	5,663
General fund, market value adjustment	61,712	40,661
Unrecognized change in market value of investments	69,150	(43,373)
Cash, end of year	**$121,183**	**$ 87,860**

The accompanying notes to financial statements are an integral part of this statement.

Notes to Financial Statements

(1) Significant Accounting Policies

Investments:

The Association accounts for its investments on a market value basis. Under the method used by the Association to value investments, the change in market value of corporate stocks during the year, after adjusting for an inflation factor (9.0 percent in 1979 and 8.3 percent in 1978), is recognized in income over a three-year period. The change in market value of treasury bills, commercial paper, etc., is reflected currently in income. The changes in market value of investments are allocated to the general and restricted fund balances as appropriate.

Accrual for Directory:

Approximately every three to five years, the Association publishes a directory which lists, among other things, the names and addresses of its membership. This directory was published in 197[illegible] and distributed at no cost to the membership. In order to match more properly the publishing cost of this directory with revenue from membership dues, the Association provided $55,000 in 1979 and $50,000 in 1978 for estimated publishing costs which will reduce actual directory expense in the year of publication.

Deferred Income:

Revenue from membership dues and subscriptions to the various periodicals of the Association are deferred when received; these amounts are then recognized as income as publications are mailed to the members and subscribers.

Revenue from life membership dues is recognized over the estimated average life of these members

(2) Investments and Investment Income

The following is a summary of investments held by the Association at December 31:

	1979		1978	
	Cost	Market	Cost	Market
Treasury bills, commercial paper, etc.—	$ 602,106	$ 602,106	$ 623,068	$ 623,068
Corporate stocks	678,589	939,270	586,968	710,117
	$1,280,695	$1,541,376	$1,210,036	$1,333,185

Investment gains (losses) recognized in income for the years ended December 31, were as follows:

	1979	1978
Treasury bills, commercial paper, etc.		
Interest	$50,722	$38,968
Change in market value	–	–
	50,722	38,968
Corporate stocks—		
Cash dividends	32,096	21,732
Increase (decline) in market value recognized (Note 3)	(4,155)	(21,530)
	27,941	202
Less investment gains allocated to restricted fund (Note 4)	7,275	867
Investment gains included in income	$71,388	$38,303

(3) Unrecognized Change in Market Value of Investments

As described more fully in Note 1, the Association recognizes in income over a three-year period changes in the market value of its corporate stocks. The following summarizes the years in which market value changes in stocks occurred that affect 1979 and 1978 revenues, and the amount of these market value increases (declines) that will be recognized in income in future periods.

Year of Market Value Change	Recognized in Income in		To be Recognized in		Unrecognized Change December 31	
	1979	1978	1980	1981	1979	1978
1976	$ –	$ 8,100	$ –	$ –	$ –	$ –
1977	(15,461)	(15,461)	–	–	–	(15,461)
1978	(14,169)	(14,169)	(14,168)	–	(14,168)	(28,337)
1979	25,475	–	25,474	25,475	50,949	–
	$ (4,155)	$(21,530)	$11,306	$25,475	$36,781	$(43,798)

Included in the above unrecognized changes as of December 31, 1978, is a decline of $2,584 which has been allocated to a restricted fund. The amount allocated is based on the percentage of the Association's total stock portfolio owned by this restricted fund. During 1979 the portion of the total stock portfolio owned by this restricted fund was liquidated.

(4) Restricted Fund:

The Association sponsors the Economics Institute, an organization that provides orientation programs for foreign graduate students of economics. The Policy and Advisory Board which determines overall policies applicable to Economics Institute is appointed by the President of the Association. In October 1979, the Economics Institute incorporated as an entity distinct from the Association. Prior to October 1979, the Economics Institute participated in the investment program of the Association and its share of investments were accounted for as restricted funds in the accompanying statement of assets and liabilities. In October 1979, the Economics Institute liquidated its share of the Association's investments. Investment income and market value adjustments applicable to the Economics Institute which wereallocated to the restricted fund prior to liquidation were as follows:

	1979	1978
Net investment gains (Note 2)	$7,275	$ 867
Market value adjustments arising from inflation	2,199	2,549
	$9,474	$3,416

(5) Retirement Annuity Plan

Employees of the Association are eligible for participation in a contributory retirement annuity plan. Payments by the Association and participating employees are based on the employee's compensation. Benefit payments are based on the amounts accumulated from such contributions. The total pension expense was $16,436 and $14,762 for 1979 and 1978, respectively.

(6) The Association

The American Economic Association files its federal income tax return as an educational organization, substantially exempt from income tax under section 501(c)(3) of the U.S. Internal Revenue Code. As required by Section 511(a) of this Code, the Association provides for federal income taxes on certain revenues which are not substantially related to its tax exempt purpose. This "unrelated business income" includes income from advertising and the sale of mailing lists.

The Association has been determined to be an organization which is not a private foundation.

EXHIBIT 1—THE AMERICAN ECONOMIC ASSOCIATION STATEMENT OF OTHER GENERAL AND ADMINISTRATIVE EXPENSES FOR THE YEARS ENDED DECEMBER 31, 1979 AND 1978

	1979	1978
Mailing list file maintenance and periodic mailing expenses	$28,201	$ 37,563
Accounting and legal	9,600	9,200
Office supplies	7,659	11,102
Postage	14,210	15,497
Dues and subscriptions	5,895	3,405
Telephone	2,543	3,260
Investment counsel and custodian fees	4,499	3,513
President and president-elect expenses	3,985	4,490
Travel and entertainment	726	858
Depreciation (straight-line method)	1,295	1,231
Uncollectible receivables	507	5,822
Currency exchange charges	1,214	1,504
Insurance and miscellaneous	3,514	5,702
	$83,848	$103,147

Report of the Managing Editor
American Economic Review

We received 719 papers in 1979 and printed 119. The number received is somewhat larger than that in each of the last three years, but still below the all-time high of 879 received in 1970. The history of papers submitted and published is contained in Table 1. There is presently a backlog of 84 accepted papers; 29 will appear in March 1980, and the remainder in June or September.

TABLE 1—MANUSCRIPTS SUBMITTED AND PUBLISHED, 1960–79

Year	Submitted	Published	Ratio of Published to Submitted
1960	276	46	.17
1961	305	47	.15
1962	273	46	.17
1963	329	46	.14
1964	431	67	.16
1965	420	59	.14
1966	451	62	.14
1967	534	94	.18
1968	637	93	.15
1969	758	121	.16
1970	879	120	.14
1971	813	115	.14
1972	714	143	.20
1973	758	111	.15
1974	723	125	.17
1975	742	112	.15
1976	695	117	.17
1977	690	114	.17
1978	649	108	.17
1979	719	119	.17

The backlog of unprocessed manuscripts has been kept to six months. As of December 15, nearly all manuscripts received before the preceding June 15 have been processed. The authors have received some type of decision. It is not efficient to cut this delay time further.

This coming year will be my last as managing editor of the *Review*. My term will expire December 31, 1980. Letters and manuscripts should be addressed to my office until the end of 1980. When the new editor is appointed, the transfer of editorial responsibilities and physical facilities will be arranged, and the details announced. I will also provide my successor with full advice on how to avoid mistakes of the past. A transition of this kind will impose strains on expectant authors, and I hope they will bear with us while we make the change. The size of the *Review* will be somewhat larger in 1980, in order to limit the backlog of accepted manuscripts bequeathed to my successor.

Number and Subject Matter of Printed Papers

As Table 2 shows, we printed 119 papers this year, 52 main articles and 67 shorter papers consisting of communications, notes, comments, and replies. The size of the *Review* is slightly larger than last year, 1,058 pages in 1979, compared to 1,035 in 1978 and 1,067 in 1977.

TABLE 2—SUMMARY OF CONTENTS, 1978 AND 1979

	1978		1979	
	Number	Pages	Number	Pages
Articles	52	663	52	635
Shorter Papers, including Notes, Comments and Replies	56	299	67	353
Special Articles	1	2		
Dissertations		24		26
Announcements and Notes		37		35
Index		10		9
TOTAL	109	1035	119	1058

TABLE 3—SUBJECT MATTER DISTRIBUTION OF SUBMITTED AND PUBLISHED MANUSCRIPTS IN 1979

	Submitted	Published
General Economics and General Equilibrium Theory	18	4
Micro-Economic Theory	157	23
Macro-Economic Theory	69	3
Welfare Theory and Social Choice	31	12
Economic History, History of Thought, Methodology	17	4
Economic Systems	24	1
Economic Growth, Development, Planning, Fluctuations	22	5
Economic Statistics and Quantitative Methods	19	8
Monetary and Financial Theory and Institutions	46	9
Fiscal Policy and Public Finance	35	5
International Economics	82	16
Administration, Business Finance	12	2
Industrial Organization	46	5
Agriculture, Natural Resources	19	4
Manpower, Labor Population	92	11
Welfare Programs, Consumer Economics, Urban and Regional Economics	30	7
TOTAL	719	119

Table 3 shows the distribution of submitted and published manuscripts, classified by subject matter. The most popular fields continue to be microeconomics, labor, and international economics.

Expenses—Printing and Mailing

Tables 4A and 4B show the printing and mailing expenses for the four regular issues and the *Papers and Proceedings* issue of the *Review*, for 1978 and 1979. Table 4A is a revision of the Table 4 printed in my annual report for 1978 (May 1979, p. 405). Two corrections have been made: The December 1978 costs are actual rather than estimated. Second, to conform with a recommendation by the Association's auditors, we now list the expenses of producing reprints ($8,366 in 1978), but no longer take a credit for the charges billed to authors. This increases our costs as reported but leaves the Association's cash position unaffected.

As Table 4B shows, total printing and mailing expenses came to $222,405 in 1979, with roughly 25 percent going to the *Papers and Proceedings* issue.

I expect the printing and mailing charges to increase in 1980, but the amount projected is only a crude guess. Mailing charges and paper costs are likely to go up. However we have switched to a new vendor of typesetting service, a factor that should produce some saving in cost. The change begins with the March 1980 issue, and I expect to save money on galley and pageproof corrections. The new vendor appears to be more competent and reliable than his predecessor, and this should lead to fewer errors.

In summary, I project our 1980 expenditures to reach $350,000. Expenses for earlier years are shown in Table 5.

Papers and Proceedings

Our second volume of *Papers and Proceedings* appeared in May 1979. Prior to the

TABLE 4A—COPIES PRINTED, SIZE, AND COST OF PRINTING AND MAILING 1978 (Revised) *AER*

Issues:	Copies Printed	Pages Net	Pages Gross	Costs: Issue	Costs: Reprints	Costs: Total
March	27,463	245	288	$32,320.68	$1,117.79	$33,438.47
May	27,864	520	536	55,979.64	2,806.29	58,785.93
June	28,000	265	288	35,205.02	1,684.51	36,889.53
September	28,000	244	280	33,925.61	1,241.47	35,167.08
December	27,500	298	336	44,756.92	1,516.18	46,273.10
Annual Misc.						2,987.33[a]
TOTAL	138,827	1,572	1,728	$202,187.87	$8,366.24	$213,541.44

TABLE 4B—COPIES PRINTED, SIZE, AND COST OF PRINTING AND MAILING 1979 *AER*

March	28,000	248	288	$35,493.75	$1,360.04	$36,853.79
May	27,500	435	448	53,659.98	2,684.37	56,344.35
June	27,500	247	268	34,219.16	1,406.40	35,625.56
September	28,000	274	312	39,630.81	1,478.64	41,109.45
December	27,500	298	352	48,717.76	1,435.84	50,153.60
Annual Misc.						2,318.26[a]
TOTAL	138,500	1,502	1,668	$211,721.46	$8,365.29	$222,405.01

[a]Includes costs of addressing and storage; express charges.

TABLE 5—ACTUAL AND BUDGETED EXPENDITURES, 1972–80

	Printing and Mailing	Office Expenses	Total
1972	$107,196	$44,473	$151,669
1973	117,873	49,121	166,994
1974	139,502	58,396	197,898
1975	129,476	63,372	192,848
1976	139,300	67,130	206,430
1977	141,769	70,788	212,557
1978	213,541[c]	89,099	302,640
1979[a]	207,635[c]	97,564	305,199
1979[b]	222,405[c]	92,218	314,623
1980[a]	239,593[c]	111,754	351,347

[a]Budget.
[b]Actual.
[c]Includes *Papers and Proceedings*.

May 1978 issue, this volume had been the responsibility of the Secretary's office. Thanks to the hard work of Wilma St. John and James Hanson, this second volume was turned out with fewer problems than the first. We were also aided by the earlier date of the annual meeting in 1978, giving us more time to produce the work. With the biennial scheduling of summer meetings, we will again need an intensive effort in 1980 to produce a volume for May, and then turn to manuscripts that will be presented to the September meeting. Because my colleague James Hanson is on leave of absence this year, I have asked Daniel Spulber to assist me with the May 1980 issue.

The Board of Editors

The Board consists of eighteen members, chosen by the managing editor, with the approval of the Executive Committee of the Association. Their names are printed on the contents page in every issue. The Board is responsible for two of the most difficult parts of the refereeing process: acting as super referees in the case of complaints over the fairness, objectivity, or competence of other referees; and determining the quality of comments on published articles. Their cooperation is very satisfactory.

The Board also advises the managing editor on the conduct of his job, criticizes poor performance and poor articles, and acts as a professional conscience. My interaction with the Board has been very helpful.

In March 1979, six new members of the Board were appointed by the Executive Committee for three-year terms. They are

Pranab Bardhan, Peter Diamond, W. Erwin Diewert, Michael Parkin, Roy Radner, and Nancy Schwartz.

Six members of the Board will complete their terms at the end of 1979: Albert Ando, Elizabeth Bailey, David Bradford, Martin Feldstein, F. M. Scherer, and Jerome Stein. I wish to thank them for their high professional standards, work and cooperation. I also wish to acknowledge with thanks the services of the continuing members: Rudiger Dornbusch, William Oakland, Richard Roll, Michael Spence, William Vickrey, and S. Y. Wu.

Acknowledgments

I should like to thank my associates for their devotion and hard work: Wilma St. John for her fine work as assistant editor on the regular issues of the *Review* plus the *Papers and Proceedings*; my colleague James Hanson who served as co-editor of the *Papers and Proceedings*, Deborah Franklin our editorial assistant, and Sandra Overton our secretary.

The following graduate students have worked for the *Review* this year as proofreaders and hunters of false proofs: John Boschen, George Briden, John Chilton, Keehwan Park and Joel Scheraga.

The following economists have served as editorial consultants in the screening of manuscripts:

W. Adams
J. Albrecht
F. E. Bloch
L. Blume
R. Bolton
G. J. Borjas
K. Boyer
R. P. Braeutigam
A. T. Denzau
G. Dorman
P. Evans
R. Falvey
R. Feldman
M. T. Flaherty
R. E. Franck, Jr.
D. E. Frey
H. L. Gabel
J. Geweke
G. Goldstein
F. Gollop
E. J. Green
D. Hanson
M. Hashimoto
R. Hendrick
C. Lieberman
R. J. Mackay
N. P. Obst
L. Papademos
J. D. Richardson
R. Riezman
J. Roberts
H. Rosen
M. Rosenzweig
T. Russell
J. Rutledge
A. Schotter
S. Shavell
M. Simon
C. F. Stone
A. Strickland
J. Trapani
B. Wasow
L. Wilde
R. Wilder
K. Wolpin
D. C. Zannoni
A. Zelenitz

In addition to the members of the Board and the editorial consultants, I have sought and received the assistance of a large number of economists during the year. I wish to thank them for their cooperation and high standards in reading and evaluating manuscripts. The following have assisted as referees:

H. Aaron
B. Aghevli
G. Akerlof
J. S. Akin
A. Alchian
J. Anderson
J. Artus
O. Ashenfelter
A. J. Auerbach
C. Azariadis
C. Azzi
M. Bailey
E. Baltensperger
A. Barnett
D. Baron
N. S. Barrett
R. J. Barro
J. Barron
A. Bartel
W. Baumol
Y. Barzel
R. N. Batra
W. Becker
M. Beckmann
E. Berglas
B. Bergmann
T. Bergstrom
B. Bernanke
E. Berndt
G. O. Bierwag
D. Bigman
J. Bilson
H. P. Binswanger
M. Blejer
R. G. Bodkin
J. P. Bonin
C. F. J. Boonekamp
T. E. Borcherding
G. Borjas
M. Boskin
R. Boyer
W. H. Branson
F. R. P. Brechling
E. K. Browning
J. Buchanan
K. Burdett
H. S. Burness
D. Capozza
C. D. Campbell
J. A. Carlson
D. Carlton
D. Caves
W. Chang
P. L. Cheng
C. F. Christ
L. Christensen
C. Clark
C. Clotfelter
R. Coen
D. V. Coes
B. C. Conley
J. Conlisk
M. Connolly
P. J. Cook
R. Cooter
W. M. Corden
J. C. Cox
M. A. Crew
M. L. Cropper
A. Cukierman

M. R. Darby
S. Das
P. Dasgupta
P. David
E. Davis
R. Deacon
S. De Canio
H. Demsetz
M. Denny
A. De Vany
D. Dewey
P. Doeringer
F. T. Dolbear, Jr.
W. Dolde
R. Dorfman
L. Dudley
J. H. Duloy
J. C. Dyer IV
C. Eaton
R. Ehrenberg
I. Ehrlich
R. Eisner
S. Ekern
B. Ellickson
E. Elton
W. Enders
D. Epple
T. W. Epps
R. Evenson
R. Falvey
E. Fama
G. Feder
G. Feiger
A. Feldman
R. Findlay
P. C. Fishburn
A. Fisher
F. Fisher
A. Fishlow
R. Flanagan
B. Fleisher
B. Friedman
D. Friedman
I. Friend
G. Fromm
E. Furubotn
M. Fuss
R. Gallman
B. Gardner
A. Gifford
L. Girton
S. M. Goldfeld
R. J. Gordon
J. A. Gray
E. Green
E. Greenberg
M. L. Greenhut
R. Grieson
J. Griffin
Z. Griliches
H. Grossman
S. Grossman
H. Grubel
M. Guitian
A. Gustman
J. Gwartney
R. Hamada
M. Hamburger
D. Hamermesh
B. W. Hamilton
W. L. Hansen
J. R. Haring
J. P. Harkness
M. Harris
O. Hart
R. Hartman
J. M. Hartwick
J. Hausman
G. A. Hawawini
J. Heckman
J. Hekman
M. Hellwig
H. Herberg
H. Hochman
W. Holahan
C. C. Holt
D. Holthausen
T. Horst
P. Howitt
E. P. Howrey
C. Hulten
Y. M. Ioannides
P. Isard
N. L. Jacob
D. Jaffee
L. Johansen
G. Johnson
L. Johnson
R. Jones
M. Jones-Lee
P. Joskow
R. Just
H. B. Kaitz
E. Kalachek
M. Kamien
E. J. Kane
Y. Kanemoto
D. Karnosky
T. Kennedy
L. Klein
R. Klein
B. P. Klotz
L. Kotlikoff
Y. Kotowitz
I. B. Kravis
A. Krueger
D. Laidler
K. J. Lancaster
H. E. Lapan
L. B. Lave
E. Lazear
D. Leigh
H. Leland
S. LeRoy
H. Levy
K. Lewis
T. Lewis
D. Lilien
L. A. Lillard
C. Link
S. F. Lippman
J. Little
R. Litzenberger
P. J. Lloyd
N. V. Long
M. C. Lovell
R. A. McCain
J. J. McCall
B. T. McCallum
T. McGuire
L. McKenzie
R. McKinnon
R. J. Mackay
W. Magat
J. H. Makin
B. Malkiel
A. L. Marty
T. Mayer
W. Mayer
D. Mayers
J. Mayshar
J. Medoff
M. Mendelson
R. T. Michael
P. Mieszkowski
M. A. Miles
D. E. Mills
J. J. Minarik
F. Mishkin
B. Mitchell
T. Miyao
H. Mohring
J. Muellbauer
G. Munro
P. Musgrove
R. Muth
R. Myerson
K. Nagatani
A. Nakamura
P. Neher
J. Nelson
D. M. G. Newbery
J. Niehans
Y. Niho
R. Noll
W. Nordhaus
W. E. Oates
R. Oaxaca
J. Ochs
L. H. Officer
E. Olsen
M. Olson
J. A. Ordover
L. Orren
D. K. Osborne
B. Owen
J. Panzar
J. Paroush
D. Parsons
M. V. Pauly
J. Peles
S. Peltzman
J. Pelzman
M. Perry
J. Pettengill
N. Phelps
R. S. Pindyck
J. E. Pippenger
C. G. Plourde
H. M. Polemarchakis
M. Polinsky
W. Poole
R. Porter
R. A. Posner
F. L. Pryor
B. Putnam
R. H. Rasche
A. Raviv

A. Razin
S. Resnick
V. V. Roley
S. Rosen
S. Rosenfielde
J. Rosse
M. Rothschild
S. Rottenberg
D. Rubinfeld
M. Rubinstein
R. J. Ruffin
W. R. Russell
J. Rutledge
H. E. Ryder, Jr.
E. Sadka
M. Salemi
D. Salkever
S. Salop
A. Sandmo
A. Santomero
T. Sargent
K. Sato
R. Sato
T. Saving
L. D. Schall
B. Schiller
R. Schmalensee
A. Schmitz
A. Schotter
R. Schuler
T. P. Schultz
W. Schulze
M. Schupack
A. J. Schwartz
D. Schwartzman
A. Schweinberger
S. Schwert
G. W. Scully
J. Seater
E. Seskin
S. Shavell
R. Sherman
E. Sheshinski
R. J. Shiller
R. Shishko
J. Shoven
J. Siegel
W. Silber
E. Silberberg
D. Sjoquist
D. Small
J. P. Smith
J. J. Soladay
R. Soligo
D. Spulber
F. Stafford
D. A. Starrett
N. Stern
R. M. Stern
G. Stevens
R. Stevenson
D. B. Stewart
M. Stewart
G. Sweeney
P. J. Taubman
M. Taussig
J. Taylor
E. Thorbecke
R. Toikka
R. D. Tollison
E. Tower
R. Townsend
S. Turnovsky
D. Usher
A. Vanags
H. R. Varian
D. R. Vining, Jr.
M. Visscher
W. Vroman
P. Wachtel
M. L. Wachter
T. J. Wales
F. Warren-Boulton
H. W. Watts
R. Waud
L. Wegge
L. Weiss
L. W. Weiss
Y. Weiss
M. L. Weitzman
F. Welch
F. W. Westfield
W. C. Wheaton
L. J. White
M. D. White
D. Wildasin
T. D. Willett
J. Williams
R. Willig
R. Wilson
D. Wise
K. Wolpin
G. Wood
J. H. Wood
A. D. Woodland
G. Yaniv
J. Yawitz
W. P. Yohe
L. Young
R. Zeckhauser

GEORGE H. BORTS, *Managing Editor*

Report of the Managing Editor

Journal of Economic Literature

This year, 1979, has been (*thus far*, any prudent editor would add) routine and uneventful.

Members will note (particularly after they have received the December issue which should be waiting for them when they get home from this convention) that we have published four survey articles, about a dozen articles reacting to literature or reacting to the reactions (as in the case of Sir John Hicks on Alan Coddington), and one article on a variant literary form, the filmstrip.

Tables 1 and 2 illustrate the allocation of space in the *Journal of Economic Literature* for 1979 as well as comparisons for the years 1969 through 1978 and, finally, Table 3 classifies material by technical difficulty.

During 1979 we published the annual *Index* for 1975. During 1980 we plan to produce the annual *Indexes* for 1976 and 1977.

Looking to 1980, there will be four survey articles. One will be on the "Labor Union Effects on Wage Gains"; another will be on the "Measurement of the Impact of Technical Progress on the Demand for Intermediate Goods," and one is almost certain to be on, "Research, Technical Change and Economic Analysis." Three others are in the mill, but only one of them will be used in 1980.

In terms of essays on the literature, there will be two essays on industrial organization: "The Corporation and the Neoclassical Paradigm" and "Industrial Organization, Corporate Strategy and Structure." In addition we plan to have two articles focusing on the contributions of two seminal scholars, Sir Dennis Robertson and Harold Hotelling. There will be at least two other articles on leading scholars' reactions to contemporary literature: one will be on "Inflation and Taxation" and the other will be on "Macroeconomic Literature and the Uses to Which it can be Put in Developing Countries." Beyond that, there will be a short piece on productivity in the pin manufacturing industry since 1776 and an

TABLE 1—QUANTITATIVE ANALYSIS OF *JEL* CONTENTS, 1975–79
(Number of pages in parentheses)

	1975		1976		1977		1978		1979	
	No.	Pages	No.	Pages	No.	Pages	No.	Pages	No.	Pages
Survey articles	3	(119)	3	(116)	4	(127)	4	(180)	4	(175)
Essays on subfields	5	(100)	8	(157)	4	(99)	4	(79)	8	(147)
Review articles	–	–	–	–	1	(9)	1	(12)	–	–
Articles about economic literature	1	(11)	–	–	–	–	2	(39)	1	(15)
Communications	12	(36)	2	(7)	15	(62)	1	(3)	–	–
Books annotated	1203	(223)	1204	(253)	1212	(246)	1200	(259)	1201	(258)
Books reviewed	183	(282)	185	(278)	172	(274)	182	(286)	166	(293)
Journal issues listed and indexed	908	(177)	901	(159)	970	(174)	921	(180)	962	(187)
Number of individual articles	6788	–	6211	–	7164	–	7344	–	7437	–
Subject index of journal articles	–	(349)	–	(328)	–	(329)	–	(360)	–	(377)
Abstracts of articles	1637	(331)	1502	(309)	1589	(326)	1649	(338)	1645	(336)
Total pages[a]		(1700)		(1644)		(1713)		(1873)		(1877)

[a]Includes, in addition to listed pages, classification systems, table of contents, indices, journal subscription information, etc.

TABLE 2—CLASSIFICATION BY SUBJECT, 1969–79

		1979		1969–79
		Commissioned Surveys	Creative Curmudgeon Essays	All Articles Total[a]
01	General	1	1	10
02	Theory	3	4	33
03	Thought (Methodology)	–	3	26
04	Economic History	–	–	4
05	Comparative Systems	–	–	4
11–12	Growth and Development	–	1	7
13	Stabilization	–	–	2
21–22	Econometric, Statistical Theory, Statistics	–	–	3
31	Monetary Economics	–	–	7
32	Fiscal Economics	–	–	6
40–44	International Economics	–	–	11
50	Managerial Economics	–	–	1
60	Industrial Organization, Industrial Regulation	–	–	1
70	Agricultural and Resource Economics	–	–	2
80	Labor Economics	–	–	8
90	Applied Welfare Economics, Regional Economics	–	–	7
	TOTALS	4	7	132

[a]Includes all review articles on books, general essays on all literature.

TABLE 3—CLASSIFICATION BY TECHNICAL DIFFICULTY, 1969–79

	1979		1969–79: Totals: Surveys; Creative
	Surveys	Creative Curmudgeon Articles	Curmudgeon Articles; Others[a]
Most Difficult	2	–	23
Some Difficulty	1	3	57
Not Difficult	1	6	52
TOTALS	4	9	132

[a]Review articles or books and general essays on all literature; excludes very short communications.

exchange of views on Walrasian economics between two leading scholars, and at least one review article exploring the reasons for the lack of impact on the economics profession of Georg Simmel's book, *Philosopy of Money*.

The Chancellor (Dr. Wesley Posvar), the Provost (Dr. Rhoten Smith), the Dean of the Faculty of Arts and Sciences (Dr. Jerome Rosenberg), and particularly the Senior Vice Chancellor for Operations (Dr. Jack Freeman) of the University of Pittsburgh have again allocated University of Pittsburgh support to the *Journal*. Their willingness to do so for 1980, even though it had originally been assumed that 1979 would be the *Journal*'s last year at Pitt, illustrates an understanding of and an enthusiasm for scholarly work in the economics discipline. I thank them frequently privately, and I take this opportunity again to do so publicly.

Four members of the Board of Editors have now completed their terms. I wish to convey to them publicity my great appreciation for their tremendous help. They are: Solomon Fabricant, William H. Miernyk, Michael J. Piore, and Barbara B. Reagan. The other members of my Board who will continue for one or two more years have been similarly cooperative: I look forward to our continuing associations. I have nominated several people to replace those whose terms have been completed.

I also wish to thank the following economists (plus several who have chosen to remain anonymous) for advice and assistance in the commissioning, refereeing, and revising of articles.

M. Abramovitz	I. B. Kravis
A. K. Ando	R. E. Lucas, Jr.
W. Baer	F. Machlup
R. J. Barro	T. Marschak
W. Baumol	A. H. Meltzer
M. Bronfenbrenner	N. Miller
J. M. Buchanan	J. Niehans
P. Deane	H. C. Recktenwald
R. Fels	J. Robinson
D. R. Fusfeld	P. A. Samuelson
W. D. Grampp	P. Saunders
J. Hicks	M. Shubik
M. D. Intriligator	J. L. Simon
B. Kantor	O. E. Williamson

The practice in this *Journal* has been more and more frequently to reveal to authors who their referees have been. Perhaps because virtually everything in this *Journal* is commissioned, and because most of the writers are not *ingenus*, the practice generally works out satisfactorily. Authors should be professionals able to accept and reject referees' comments; the referees should be careful in their comments and should not seek to substitute willy-nilly their own judgments for an author's views. We have lived in a period when it has been assumed that many editors make their assessments on extraneous considerations, particularly the authors' affiliations; my own considered view is that those who believe that do a disservice both to referees and to authors. Critics should know that their criticisms will be taken seriously; authors often are aided if they realize that they have failed to persuade people in whose judgment they have the greatest confidence. I mention this in this annual report because I think that the experience in the *Journal of Economic Literature* offers an alternative assessment to the "blind refereeing" of articles by "anonymous" referees.

Finally, each of the members of the *Journal* staff has done his (her) usual superlative work during this year; the Associate Editor, Naomi Perlman; the Assistant Editor, Drucilla Ekwurzel; the principal secretary, Lyndis Rankin; the clerk for the quarterly and annual indexing, Margaret Yanchosek; my administrative assistant (who works half-time for this *Journal*), Eileen Fiedler; and my consultant on substantive indexing decisions,, Asatoshi Maeshiro (University of Pittsburgh).

MARK PERLMAN, *Managing Editor*

Report of the Director
Job Openings For Economists

During 1979, employers advertised 1,928 new vacancies, a record number. Of these 1,289 (67 percent) were classified as academic and 639 (33 percent) were nonacademic. Last year employers advertised 1,647 new vacancies; 70 percent were academic and 30 percent nonacademic. This division between academic and nonacademic of roughly two-to-one has been the same for several years. Table 1 shows total listings (employers), total vacancies, new listings and new vacancies by type for each issue of *JOE* in 1979.

TABLE 1—JOB LISTINGS FOR 1979

Issue	Total Listings	Total Jobs	New Listings	New Jobs
Academic				
February	115	259	85	169
April	70	128	66	120
June	43	61	37	55
August	53	129	48	121
October	124	285	116	268
November	106	238	106	238
December	220	555	113	318
Subtotals	731	1,655	571	1,289
Nonacademic				
February	22	131	19	116
April	26	84	22	64
June	29	112	24	91
August	26	72	19	45
October	52	207	46	161
November	18	63	18	63
December	42	188	27	99
Subtotals	215	857	175	639
TOTALS	946	2,512	746	1,928

Universities with graduate programs and four-year colleges continue to be the major sources of job listings. They constitute 43 and 34 percent, respectively, of total employers. This compares to last year's 44 and 37 percent for the two. Table 2 shows the number of employers by type for each 1979 issue. The distribution is similar to that in 1976, 1977, and 1978.

The field of specialization most in demand continues to be general economic theory. Generalists with a strong background in mathematics and statistics appear to be the type of economist that employers are seeking. The applied area of specialization seems to be of secondary importance. Table 3 shows the number of citations by field of specialization. General economic theory (000) led, followed by monetary and fiscal (300), econometrics and statistics (200), and welfare and urban (900). This pattern is also the same as that of the past several years.

The proposed 1980 budget and the 1979 (adopted and estimated) and 1978 (adopted and actual) budgets are given in Table 4. The 1979 approved budget projected a deficit (including indirect costs) of \$10 thousand. The estimated actual deficit is \$10.2 thousand. Total revenues are expected to be \$19.4 thousand, total direct costs \$11.1 thousand, and total indirect costs \$18.5 thousand. The proposed budget for 1980

TABLE 2—NUMBER AND TYPES OF EMPLOYERS LISTING POSITIONS IN *JOE* DURING 1979

Issue	Four-Year Colleges	Universities with Graduate Programs	Junior Colleges	Federal Government	State/Local Government	Banking of Finance	Business or Industry	Consulting or Research	Other	Total
February	59	55	1	8	1	2	1	7	3	137
April	39	31	–	6	3	2	2	8	5	96
June	22	2n21	–	8	4	3	1	9	4	72
August	19	34	–	6	4	4	5	5	2	99
October	49	75	–	15	1	4	4	21	7	176
November	40	66	–	7	1	1	2	5	2	124
December	98	122	–	19	2	–	1	18	2	262
TOTALS	326	404	1	69	16	16	16	73	25	946

TABLE 3—FIELDS OF SPECIALIZATION CITED: 1979

Field[a]	February	April	June	August	October	November	December	Totals
General Economic Theory (000)	106	65	55	53	145	87	257	768
Growth and Development (100)	33	23	22	17	47	21	52	215
Econometrics and Statistics (200)	56	45	32	24	61	39	103	360
Monetary and Fiscal (300)	57	30	30	37	73	45	138	410
International Economics (400)	26	19	16	21	47	27	59	215
Business Administration, Finance, Marketing and Accounting (500)	61	39	21	15	50	24	64	274
Industrial Organization (600)	39	33	25	19	53	35	71	275
Agriculture and Natural Resources (700)	21	21	16	21	35	18	48	180
Labor (800)	26	20	11	10	28	21	52	168
Welfare and Urban (900)	40	23	19	18	63	37	82	282
Related Disciplines (A00)	8	3	3	2	6	8	15	45
Administrative Positions (B00)	9	10	1	2	12	11	20	65
TOTALS	482	331	251	239	620	373	961	3,257

[a]Fields of specialization codes are from the *Journal of Economic Literature*.

TABLE 4—JOB OPENING FOR ECONOMISTS
Budget for 1980 (in thousands)

	1977		1978		1979		1980
	(Adopted)	(Actual)	(Adopted)	(Actual)	(Adopted)	(Estimated)	(Proposed)
Revenue:							
Subscriptions		19.3		16.0		19.2	
Miscellaneous		.7		.3		.2	
Total Revenue	$21	$20	$20	$16.3	$20	$19.4	$20.0
Expenses:							
Direct:							
Computer		.8		1.5	2	.4	.5
Typewriter Rental		2.4		.8	2	1.3	2
Postage		3.4		3.9	4	4.4	5
Printing		3.6		4.4	4	5	5
Salaries		.2					
Miscellaneous		.5		.4	1		.5
Total Direct		10.9		10.9	14	11.1	13
Indirect:							
Salaries		16.1		17.5	16	18.5	22
Other							
Total Indirect		16.1		17.5	16	18.5	22
Total Expenses	$25	$27.0	$27	$28.4	$30	$29.6	$35
SURPLUS (DEFICIT)	(4)	(7)	(7)	(12.1)	(10)	(10.2)	(15)

projects revenues of $20 thousand, total direct costs of $13 thousand, and total indirect costs of $22 thousand. This leads to a projected accounting deficit of $15 thousand.

Violet Sikes continues to do virtually all the work involved in the publication and distribution of *JOE*. I wish to express my gratitude to her for a splendid job.

C. ELTON HINSHAW, *Director*

The Committee on the Status of Women in the Economics Profession

In establishing the Committee on the Status of Women in the Economics Profession (CSWEP) in 1971, the American Economic Association recognized that women were not sufficiently represented in the economics profession and gave official sanction to efforts to increase their role and participation in economics. To this end, CSWEP has undertaken a number of activities aimed at increasing the number of women active in the profession and has attempted to monitor their role and activities in economics. Thus this report will briefly discuss CSWEP's activities and the status of women in the academic labor market for economists.

I. CSWEP Activities

CSWEP's activities fall into two major groups: those aimed at enhancing the workings of the labor market for economists, particularly with respect to women; and those aimed at increasing the visibility of women and women's issues in the economics profession.

In terms of informational activities, the maintenance of the roster and publication of the *Newsletter* are the most important. CSWEP maintains a roster of all women economists who have registered with it. Each listing on the roster states the highest degree earned, current job, and fields of interest. Thus potential employers can obtain from the roster a list of all potential women candidates who fit a given job description (for example, economists specializing in money and banking with more than three years experience). Consequently, broad use of the roster should ensure that no potential woman candidate will be excluded from consideration for lack of information. Although the roster listings are sold at a modest fee, they are currently used by a relatively small number of academic departments and government agencies. Hence one of CSWEP's activities this past year has been an analysis of roster usage to de termine how it could be used more effec tively. Ideally, no job should be fille without reference to the roster to determin if there are qualified women available. Thu a major goal of CSWEP in the coming yea is to increase the usage and effectiveness o the roster.

While the roster primarily serves th needs of potential employers to ensure tha they have full information about the pool c women candidates, the list of jobs in th *Newsletter* serves the needs of potentia candidates. Although much of the job infor mation in the *Newsletter* is also in th Association's publication of *Job Opening for Economists*, it is felt that the additiona listings in the *Newsletter* are worthwhile This is particularly true for people who ar not actively looking for a job, but migh learn of a suitable opening through th *Newsletter* listings. The *Newsletter* also pro vides information about issues of concern t women economists.

In addition to trying to improve the work ings of the job market by enhancing th flow of information, CSWEP has attempte to increase the participation of women i the economics profession at the annua meetings of the Association. To this end, has initiated discussions by the Executiv Committee of the Association concernin ways in which participation in the annua meetings could be broadened. While thes discussions are still at a preliminary stag CSWEP is hopeful that ways can be foun to increase the participation of less estab lished members of the profession in a aspects of the annual meetings.

Acting more directly, CSWEP has als attempted to increase the participation c women in the annual meetings by sponsor ing sessions at these meetings. Althoug these sessions do not exclusively focus o women's issues, they attempt to focus o topics that might be of concern to wome and in which women could be expected t

be working. In addition, since the papers in these sessions are contributed rather than invited, the CSWEP sessions provide an outlet for less-established economists, either male or female, at the annual meeting of the Association. At the annual meeting in 1979, the CSWEP session was entitled "Consequences of the Growth of the Two-Earner Family." The topic planned for the 1980 meetings will focus upon the impacts of macro-economic activity (particularly inflation) upon women and minorities.

CSWEP has recently extended its activities to the meetings of the regional associations. During the academic year 1979–80, CSWEP sessions have been held in the meetings of the Southern Economic Association, the Midwestern Economic Association, and the Eastern Economic Association. In this connection, CSWEP has established regional representatives who plan the CSWEP program at each of the regional meetings and encourage the participation of women economists at these sessions.

At each of the meetings of the economic associations, CSWEP also holds an open meeting and maintains a hospitality suite as a means of encouraging people to exchange their concerns about the role of women in the economics profession and discuss alternative ways to improve the role of women in the profession.

In this connection, during this past year CSWEP has taken an active role on behalf of those members of the Association who felt that it was inappropriate for the Association to hold its meetings in states that have not ratified the Equal Rights Amendment (ERA). Since the Association repeatedly refused to consider moving the site of the 1979 meeting from Atlanta,[1] CSWEP decided that it would be appropriate to use the 1979 meetings to focus on the issue of the ERA. To this end, it sponsored a session at the meetings entitled "Women's Place in the Labor Market: Will ERA Matter?" In addition, it sponsored an advertisement in the Atlanta *Constitution*, which listed some 900 economists who went on record in support of the ERA and held a press conference in support of the ERA at which Robert Solow, the president of the Association, spoke. Finally, it also sponsored an ERA reception, at which William Baumol, the president-elect of the Association, spoke as well as individuals active in the Georgia ERA movement. Consequently, CSWEP provided a vehicle for individual members of the Association to voice their support for ERA since the Association could not make such a stand.

II. The Role of Women in the Economics Profession

While CSWEP actively attempts to promote the participation and visibility of women in the economics profession, the status of women within the profession must ultimately depend upon the kinds of jobs and responsibilities they undertake. As a primarily academic profession, this is best measured by the distribution of women economists among various types of academic institutions and the flow of young women economists into these institutions.[2]

The data presented in this report came from the Universal Academic Questionnaire, distributed by the Association

[1]CSWEP made presentations before the Executive Committee of the Association in December 1977 and March 1978, urging it to move the December 1979 annual meeting from Atlanta. In both instances, the Executive Committee refused to vote to change the meeting site, citing as reasons: 1) the nonpolitical charter of the AEA; 2) the fact that such an act would constitute a secondary boycott; and 3) existing contractual obligations. The issue was raised again in the Association's Open Business Meeting, at Chicago in August 1978, and it was narrowly defeated by the members of the Association who were present at that time. In addition, CSWEP made a presentation before the Association's Executive Committee in March 1979, urging it to consider splitting the job meetings from the annual meetings, since many graduate students felt strongly about going to non-ERA states. However, the Executive Committee felt that such a move would be impractical.

[2]Reagan (1979) has an interesting paper arguing that women economists appear to be subject to the "revolving-door syndrome" under which they are hired at junior levels but not retained at senior levels. Strober and Reagan (1978) also discuss income differentials between male and female economists.

TABLE 1—DISTRIBUTION OF FULL-TIME FACULTY BY TYPE OF INSTITUTION, ACADEMIC YEAR, 1978–79

	Chairman's Group			Other Ph.D. Departments			Only M.A. Departments			Only B.A. Departments		
		Female			Female			Female			Female	
	Total	Number	Percent	Total	Number	Percent	Total	Number	Percent	Total	Number	Percent
Existing												
Professor	663	10	1.5	468	9	1.9	264	12	4.6	290	17	5.9
Associate	192	8	4.2	332	17	5.1	212	11	5.2	327	22	6.7
Assistant	340	39	11.5	287	32	11.2	238	21	8.8	403	46	11.4
Instructor	128	14	10.9	121	15	12.4	356	43	12.4	258	49	15.5
Other	131	16	12.2	90	16	17.8	183	23	12.6	153	55	36.0
New Hires												
Professor	11	–	–	9	–	–	4	–	–	6	1	16.7
Associate	13	–	–	18	–	–	3	1	33.3	26	1	3.9
Assistant	59	11	18.6	50	12	24.0	22	2	9.2	91	11	12.1
Instructor	30	2	6.7	60	7	11.7	29	5	17.2	111	5	4.5
Other	30	4	13.3	15	3	20.0	25	–	–	43	12	27.9
Promoted to Rank (1977–78)												
Professor	32	–	–	37	–	–	14	4	28.6	21	4	19.0
Associate	42	7	16.7	29	4	13.8	34	2	5.9	36	5	18.9
Assistant	9	2	22.2	4	–	–	4	2	50.0	18	2	11.1
Tenured at Rank (1977–78)												
Professor	5	–	–	8	1	12.5	–	–	–	–	–	–
Associate	13	1	7.7	23	2	8.7	21	2	9.5	26	2	7.7
Assistant	5	1	20.0	5	–	–	4	–	–	10	3	30.0
Not Rehired												
Professor	22	–	–	27	1	3.7	7	–	–	12	1	8.3
Associate	15	1	6.3	15	–	–	3	–	–	11	1	9.1
Assistant	31	2	6.5	24	6	25.0	19	2	10.5	61	1	1.6
Instructor	16	3	18.8	23	2	8.7	20	5	25.0	28	3	10.7
Other	4	2	50.0	14	7	50.0	17	1	5.9	22	4	18.2

to all department chairmen and tabulated by the Association. As such it provides the most comprehensive information concerning the academic labor market available. Since, however, responses are voluntary, information is often fragmentary and incomplete, and annual comparisons are difficult to make since the responding institutions vary. Hence the data provided by the Universal Academic Questionnaire are by no means complete and are occasionally contradictory. Thus we unfortunately do not have a fully accurate view of the role of women in the academic labor market. One of the CSWEP's planned activities for the coming year is to improve the quality of these data and the accuracy of the information concerning the role of women in the various economics departments.

Table 1 provides a summary of the distribution of academic jobs at the beginning of the academic year 1978–79 and the promotions that took place between this and the previous year. This table presents information in terms of four types of departments: the Chairman's group; other Ph.D departments; M.A. departments; and B.A departments. The Chairman's group consists of the sixty-four departments that focus on research and the training of Ph.D.'s in economics. In terms of stature, it is generally agreed that academic appointments at a department within the Chairman's group carry the most prestige. Thus this discussion will tend to focus upon the role of women in the Chairman's group as a bellwether for the entire economics profession. The other Ph.D. granting departments primarily focus

on undergraduate education, but also have a viable Ph.D. program. The M.A. departments similarly have a primary focus upon undergraduate education, but also have a Master's program. Finally, the B.A. departments are exclusively concerned with undergraduate teaching.

According to Table 1, the existing participation of women in the academic side of the economics profession appears to be distressingly small. Within the forty-nine departments reporting within the Chairman's group, there are only ten women who are full professors, eight who are associate professors, thirty-nine who are assistant professors, and fourteen who are instructors, respectively representing 1.5 percent of the full professors, 4.2 percent of the associate professors, 11.5 percent of the assistant professors, and 10.9 percent of the instructors. Although the percentage of women in each category is slightly higher for the other departments, the figures of the Chairman's group are representative.

Since tenured positions carry the most prestige within the profession it is useful to focus on them. In this connection, it is interesting to note that as of 1978–79 the reporting departments within the Chairman's group apparently feel that there are only ten women whose research and publication records are sufficiently strong to merit their appointments as full professor and eight women whose records are sufficiently strong to merit their appointment as associate professors. Moreover, during 1977–78, within the Chairman's group no woman was hired as a full professor, promoted to full professor, or hired as an associate professor. However, within the Chairman's group one woman was tenured at associate professor and seven women were promoted to associate professor during this period. This last figure is particularly encouraging since this represents 16.7 percent of the promotion to rank at this level and indicates that women are being promoted through the academic ladder.[3] It is also encouraging to note that within the Chairman's group 22.2 percent of the promotions to assistant professor were women and that 18.6 percent of the new hires were women. Since these figures represent a larger percentage than the percentage of available women, it indicates that the Chairman's group is presently taking strong affirmative action with respect to women at the junior levels.

The situation with respect to the other departments is quite similar. Thus in 1978–79, the proportion of women in tenured or tenure track positions was distressingly small. Moreover, during this year only one woman was hired as a full professor by the B.A. departments and only one woman was hired as an associate professor by each of the M.A. departments and B.A. departments. On a more encouraging note, however, a total of eight women and seven women were, respectively, promoted to full professor and associate professor among the M.A. and B.A. departments during the previous academic year. Since only one woman left the ranks of professor and associate professor within the B.A. departments, this indicates a substantial net growth in the ranks of women faculty at the senior academic levels.

In terms of changes that are occurring within the academic labor market, it is useful to consider the previous activity of those who were newly hired and the present activity of those who were not rehired. Table 2 indicates that within the Chairman's group, the M.A. departments, and the B.A. departments, relatively fewer of the newly hired women were faculty at other institutions than their male counterparts. Moreover, with the exception of the Chairman's group,

[3]Some apparent contradictions should be noted with respect to the data on promotion to associate professor. The Universal Academic Questionnaire indicated that there were, respectively, six and eight women associate professors in the Chairman's group during 1977–78 and 1978–79. However, the 1978–79 questionnaire also indicated that seven women were promoted to associate professor and that one was not rehired within this group. Hence there is a discrepancy of four women that cannot be accounted for. While part of this discrepancy may be due to differences in reporting institutions, part of it may be due to reporting errors. Given the small number of women economists in the academic labor market, it is important that their activities be recorded as accurately as possible. Hence one of CSWEP's goals is to improve the accuracy of the information in the Universal Academic Questionnaire.

TABLE 2—PREVIOUS ACTIVITY OF NEW HIRES AND CURRENT ACTIVITY OF THOSE NOT REHIRED BY TYPE OF INSTITUTION AND SEX, ACADEMIC YEAR, 1978–79

	Previous Activity of New Hires				Current Activity of Not-Rehired			
	Male		Female		Male		Female	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Chairman's Group	114	100.0	19	100.0	65	100.0	4	100.0
Faculty	43	37.7	5	26.3	32	49.2	3	75.0
Student	47	41.2	11	57.9	1	1.5	1	25.0
Government	3	2.6	1	5.3	5	7.7	–	–
Business, Banking, Research	4	3.5	2	10.5	11	16.9	–	–
Other	17	14.9	–	–	16	24.6	–	–
Other Ph.D. Departments	100	100.0	35	100.0	79	100.0	13	100.0
Faculty	32	32.0	16	45.7	29	36.7	2	15.4
Student	48	48.0	7	26.0	14	17.7	1	7.7
Government	4	4.0	1	2.9	2	2.5	–	–
Business, Banking, Research	5	5.0	1	2.9	6	7.6	3	23.1
Other	11	11.0	10	28.6	28	35.4	7	53.9
M.A. Departments	40	100.0	7	100.0	38	100.0	5	100.0
Faculty	19	47.5	3	42.9	13	34.2	1	20.0
Student	24	60.0	3	42.9	10	26.3	–	–
Government	–	–	–	–	1	2.6	–	–
Business, Banking, Research	–	–	1	14.2	4	10.5	3	60.0
Other	7	17.5	–	–	10	23.6	1	20.0
B.A. Departments	151	100.0	27	100.0	86	100.0	15	100.0
Faculty	81	53.6	12	44.5	49	57.0	6	40.0
Student	47	31.1	10	37.0	9	10.4	–	–
Government	3	2.0	–	–	1	1.2	–	–
Business, Banking, Research	9	6.0	1	3.7	9	10.4	3	20.0
Other	11	7.3	4	14.8	18	21.0	6	40.0

a significantly smaller proportion of women who were not rehired moved to other academic jobs than their male counterparts. Thus on balance, Table 2 indicates that women may have less academic mobility than men. However, the number of women involved in the groups of the newly hired and not rehired is sufficiently small to make it difficult to generalize from the available data.

On balance the information contained in Tables 1 and 2 is marginally encouraging. While the number of women in senior academic positions is distressingly small in all institutions, there does appear to be some evidence of increased numbers of promotions for women faculty members. If this is borne out by subsequent developments, it will indicate that the economics profession is undertaking substantive efforts to improve the role of women in its academic departments.

Since tenure and promotion are heavily dependent upon publication, under the sponsorship of CSWEP, Marianne Ferber and Michelle Teiman undertook an analysis of the acceptance rate of journal articles authored by male and female authors.[4] Although the number of responding journals was relatively few and hence the sample quite small, Ferber and Teiman found that the acceptance rate was significantly higher for women authors relative to male authors in journals that employed double-blind reviewing (i.e. the referee did not know

[4]Ferber and Teiman, "Are Women Economists at a Disadvantage in Publishing Journal Articles?," mimeo, December 1979.

TABLE 3—DISTRIBUTION OF SALARY FOR WOMEN FACULTY BY TYPE OF DEPARTMENT AND TIME IN RANK, ACADEMIC YEAR 1978–79

Highest Degree Offered and Relative Salary for Rank	All Women		Time in Rank			
	Number	Percent	Total	Above Median	At Median	Below Median
All Departments	90	100.0				
Salary above median	43	47.8	100.0	62.8	18.6	18.6
Salary at median	25	27.8	100.0	12.0	64.0	24.0
Salary below median	22	24.4	100.0	27.3	4.5	54.5
Ph.D., Chairman's	17	100.0				
Salary above median	8	47.1	100.0	87.5	12.5	0.0
Salary at median	6	35.3	100.0	16.7	66.7	16.7
Salary below median	5	29.4	100.0	0	20.0	80.0
Ph.D., Other	34	100.0				
Salary above median	14	41.2	100.0	78.6	7.1	14.3
Salary at median	10	29.4	100.0	10.0	70.0	20.0
Salary below median	10	29.4	100.0	40.0	0.0	60.0
M.A. Departments	23	100.0				
Salary above median	14	60.9	100.0	35.7	28.4	42.9
Salary at median	4	17.4	100.0	0.0	50.0	50.0
Salary below median	5	21.7	100.0	0.0	0.0	100.0
B.A. Departments	14	100.0				
Salary above median	7	50.0	100.0	57.1	42.9	0.0
Salary at median	5	35.7	100.0	20.0	60.0	20.0
Salary below median	2	14.3	100.0	100.0	0.0	0.0

the names of the authors and vice versa) than in journals that employed single-blind reviewing (i.e., the authors did not know the name of the referee, but the referee did know the name of the authors). While the sample is admittedly too small to be conclusive, it does indicate that discrimination may exist with respect to women authors and that their work may be discounted because they are women per se. If true, this finding has considerable negative implications for the prospects of women for promotion and tenure. CSWEP hopes that more work can be done in this important area to determine if discrimination does in fact exist with respect to women in the refereeing process. In any event, Ferber and Teiman's preliminary analysis indicates that journals should be encouraged to adopt double-blind reviewing procedures.

With respect to salary, it is important to note that the salary distribution of women faculty members appears to be in line with that of their male colleagues. This can be seen from Table 3, which gives the distribution of salary of women faculty by type of department and time in rank. In general, salary appears to be related to time in rank, with those whose time in rank is above the median having a salary that is above the median.[5] Although data are lacking to compare the distribution of male and female

[5]It should be noted, however, that the number of women faculty members considered in Table 3 is substantially less than the number of women faculty members given in Table 1. This implies that a substantial

TABLE 4—DEGREES GRANTED IN ECONOMICS BY TYPE OF DEPARTMENT AND SEX, ACADEMIC YEAR 1977–78

	All Departments	Ph.D. Departments Total	Chairman's	Other	M.A. Departments	B.A. Departments
Number Departments	373	104	49	55	57	212
Number Ph.D.s	753	753	454	299	–	–
Number female	63	63	43	20	–	–
Percent female	8.4	8.4	9.5	6.7	–	–
Number M.A.s	1,420	1,119	712	407	301	–
Number female	387	270	131	139	117	–
Percent female	27.3	24.1	18.4	34.2	38.9	
Number B.A.s	11,547	6,676	4,192	2,484	1,402	3,469
Number female	2,684	1,414	918	496	283	988
Percent female	23.2	21.2	21.9	20.0	20.2	28.5
Number Other	[a]	239	10	299	85	[a]
Number female	[a]	55	1	54	10	[a]
Percent female	[a]	63.0	10.0	23.6	11.8	[a]

[a]Not applicable.

salaries, Table 3 does not indicate the existence of any gross discrepancies between the distribution of salaries and the distribution of time in rank.

Ultimately, however, if more women are to play an active role in the economics profession, more women must be trained as economists. In this connection, Table 4 is interesting and indicates a rather sizable attrition rate between the granting of the B.A. degree and the Ph.D. degree. Specifically, in 1978–79, while 23.2 percent of all B.A. degrees were received by women, only 8.4 percent of Ph.D. degrees were received by women. Although it takes four to five years to turn a B.A. into a Ph.D., these figures are quite representative of the past four or five years[6] and clearly indicate that relatively fewer women who receive economics training at the B.A. level chose to go on to graduate school and obtain a Ph.D. than do their male counterparts. Whatever the reasons for this decline, it is clear that the flow of new female Ph.D.s must be substantially increased if the proportion of women in academic jobs is to increase substantially.

Table 5 provides information on the jobs taken by new Ph.D.s in 1978–79 and indicates that relatively fewer women Ph.D.s took academic jobs overall than their male counterparts. This was particularly true among the Chairman's group where only 44.8 percent of women Ph.D.s took academic jobs in contrast to their male counterparts of whom 65.4 percent took academic jobs. Since the Chairman's group is the primary source of Ph.D.s who take academic positions, it is somewhat discouraging to note that the majority of new women Ph.D.s from these institutions did not enter the academic labor market. Moreover, it is important to note that the total number of women who entered the academic labor market was only twenty-eight, a figure that is hardly sufficient to change the distribution of academic positions held by women. This is shown by Table 6, which indicates that only 8.9 percent of the academic jobs taken by new Ph.D.s were held by women. Moreover, within the Chairman's group only 5.0 percent of the academic jobs taken by new Ph.D.s were held by women. These figures indicate that new women Ph.D.s tend to turn away from the academic labor market. Whether this is due to perceptions of lack of opportunity is, of course, unclear.

Although relatively fewer women who receive B.A. degrees in economics go on to obtain a Ph.D. than their male counterparts, it is important to note that this is probably not due to lack of graduate student suppor

number of departments did not give the salary information on which Table 3 is based. Since it is unlikely that departments would fail to give information if women faculty member's salaries were on parity, it is likely that Table 3 overestimates the level of female faculty salaries.

[6]See CSWEP reports in *Proceedings* issues 1974, 1975, 1976, 1977, and 1978.

TABLE 5—DISTRIBUTION OF ACTIVITIES OF NEW PH.D. DEGREES BY SEX AND TYPE OF DEPARTMENT, 1978–79

	All Ph.D. Departments		Chairman's Group		Other Ph.D. Departments	
	Number	Percent	Number	Percent	Number	Percent
All Ph.D.'s						
Total	555	100.0	407	100.0	148	100.0
Education	346	62.3	260	63.9	86	58.1
Government	56	10.1	33	8.1	23	15.5
Business, Banking, Research	63	11.4	41	10.1	22	14.9
Other	90	16.2	73	17.9	17	11.5
Male Ph.D.'s						
Total	506	100.0	378	100.0	128	100.0
Education	318	62.9	247	65.4	41	55.5
Government	52	10.3	30	7.9	22	17.2
Business, Banking, Research	52	10.3	33	8.7	19	14.8
Other	84	16.5	68	18.0	16	12.5
Female Ph.D.'s						
Total	49	100.0	29	100.0	20	100.0
Education	28	57.2	13	44.8	15	75.0
Government	4	8.2	3	10.3	1	5.0
Business, Banking, Research	11	22.4	8	27.6	3	15.0
Other	6	12.2	5	17.2	1	5.0

TABLE 6—ACTIVITIES OF 1977–78 PH.D.S BY TYPE OF DEPARTMENT AND SEX

	Number of New Ph.D.s Employed in:				
	Total	Education	Government	Business, Banking, Research	Other
All Ph.D. Departments					
Total	555	346	56	63	90
Number female	49	28	4	11	6
Percent female	8.8	8.9	7.1	17.5	6.7
Chairman's group					
Total	407	260	33	41	73
Number female	29	13	3	8	5
Percent female	7.1	5.0	9.1	19.5	6.8
Other Ph.D. Departments					
Total	148	86	23	22	17
Number female	20	15	1	3	2
Percent female	13.5	17.4	4.3	13.6	11.8

on the part of the academic departments. This is shown clearly in Table 7, which indicates that although a slightly lower percentage of women students receive full support (tuition plus stipend) overall, the percentage of women students receiving some form of financial aid is virtually identical to that of male students. Thus it is likely that it is perceptions concerning their future status in the economics profession that makes women turn from graduate study in economics rather than a lack of financial support per se.

In conclusion then, although the economics profession and its related institutions have made a conscious effort to recruit and encourage women economists in recent years, it appears that, on balance, progress is still quite slow. Women economists in academic institutions still comprise an extremely small percentage of the total, and the bulk of these women hold junior level,

TABLE 7—DISTRIBUTION OF PH.D. STUDENT SUPPORT, BY TYPE OF SUPPORT, SEX, AND DEPARTMENT, 1978–79

	All Ph.D. Departments		Chairman's Group		Other Ph.D. Departments	
	Number	Percent	Number	Percent	Number	Percent
All Students						
Total	3,764	100.0	2,644	100.0	1,120	100.0
Tuition only	254	6.7	177	6.7	77	6.9
Stipend only	683	18.1	482	18.2	201	17.9
Tuition + stipend	1,500	39.9	1,028	38.9	472	42.2
No support	1,050	27.9	802	30.3	248	22.1
No record	277	7.4	155	5.9	122	10.9
Male Students						
Total	3,107	100.0	2,192	100.0	915	100.0
Tuition only	186	6.0	126	5.8	60	6.6
Stipend only	564	18.2	396	18.1	168	18.4
Tuition + stipend	1,252	40.3	865	39.5	387	42.3
No support	868	27.9	650	29.6	218	23.8
No record	237	7.6	155	7.0	82	8.9
Female Students						
Total	657	100.0	452	100.0	205	100.0
Tuition only	68	10.4	51	11.3	17	8.3
Stipent only	119	18.1	86	19.0	33	16.1
Tuition + stipend	248	37.7	163	36.1	85	41.5
No support	182	27.7	152	33.6	30	14.6
No record	40	6.1	–	–	40	19.5

nontenured positions. Although there is some evidence that young women economists are receiving promotion to the associate professor level, it is still too early to determine whether the group of able young women who entered the academic labor market in the early and mid-1970's will move up the academic ladder along with their male colleagues. Thus the real test of the commitment of the economics profession to enhance the status of women in its activities will occur in the next few years, when the presently nontenured women faculty come up for tenure and promotion. If a proportionate share of these women move up through the academic ranks, this will be a definite sign that the profession is serious about making women equal partners. If, however, a disproportionate share of young women economists are not retained, this will almost certainly be interpreted as a sign that the economics profession will remain an essentially male bastion. In this case, it is likely that able young women will increasingly turn away from economics and enter professions which they perceive will give them more attractive career opportunities.

ANN F. FRIEDLAENDER, *Chair*

Report of the Economics Institute's Policy and Advisory Board

The Economics Institute had another successful year in 1979, with enrollments for the summer program being up by 31 percent, and those for the new spring and fall sessions being up 38 and 32 percent, respectively. In total, 429 different participants enrolled in programs of training in the Institute in 1979, with their programs varying from five weeks to forty-five weeks in length, depending on individual needs. The increasing enrollment reflects a continuing and expanding need for the services of the Institute program, and continuing improvement of the program so as to provide a wider range of training opportunities for the entering foreign graduate students in economics and related fields. There is also an increased demand for the Institute program as a short-term, intensive refresher and upgrading postgraduate program for professionals from abroad.

A major and highly promising new dimension of the Institute's program is its provision of longer periods of training. This has enabled it to serve a wider range of students, particularly in terms of beginning English proficiency, and also from the point of view of providing more complete preparatory training in critical core subject areas, such as mathematics and statistics. The record suggests that many of the students who are availing themselves of this new opportunity are proving themselves to be at least as high quality students overall as the traditional Institute student who tended to have prior confirmed university placement and the associated required levels of English language proficiency before coming to the Institute. As a result, an increasing stream of excellent students are now entering graduate programs through the Economics Institute, with the Institute's record serving as an important ingredient in the placement process.

A revised edition (the sixth edition) of the Institute's *Guide to Graduate Studies in Economics and Agricultural Economics in the United States and Canada*, is being released by the Irwin Company early in the new year. The new edition considerably enlarges on the content of the previous edition, especially with respect to its analysis of program opportunities for the benefit of potential graduate students, both domestic and foreign.

Anne Krueger, University of Minnesota, and Carl Eicher, Michigan State University, completed three-year terms on the Board during the year. Douglass North, University of Washington, Dwight Perkins, Harvard University, and G. Edward Schuh, University of Minnesota, are new appointees. The Board held a regular meeting in Boulder during the summer, and a further meeting in Atlanta in December. During the year, the Institute also separately incorporated as a nonprofit educational institution.

The University of Colorado continues to serve as the Institute's host institution, although the Institute now utilizes some off-campus space as well as campus space to meet expanded needs, especially during the academic year. The Institute benefits from some continuing foundation grant support, but in the last few years, has been essentially self-supporting.

Wyn F. Owen continues as the program director.

EDWIN S. MILLS, *Chair*

Report of the Committee on U.S.–Soviet Exchanges

The fifth U.S.–Soviet Economic Symposium was held at Mount Kisco, New York, from June 10–13. The subject of the symposium was "Long-Term Structural Change in National Economies." Ten American and eleven Soviet economists participated in the discussion, which focussed on structural change in the *U.S.* and Soviet economies since 1930. The Soviet participants were quite senior people, including several academicians and directors of economic research institutes.

After the end of the symposium, the Soviet delegation spent several days each in Washington, Atlanta, Chicago, and New York. Financing for the meeting and for subsequent travel within the United States was provided by the U.S. International Communications Agency through a grant to the Association. Logistical assistance and some supplementary funding was provided by the International Research and Exchanges Board (IREX). We are grateful to these agencies for their continued support.

The Sixth Symposium is planned for Alma-Ata, U.S.S.R., at the end of May 1980, on the subject "The Role of the State in Price Formation." It is expected that ten to twelve American economists will attend, and that they will present papers in the areas of energy pricing, agricultural price policy, pricing in regulated industries, overall price stabilization, price policies of corporations with market power, and internal (transfer) pricing in large corporations.

There is continuing discussion with the Soviet exchange coordinator, Academician Tigran Khatchaturov, and with directors of Soviet economic research institutes concerning the possibility of parallel research activities by American and Soviet economists. The Soviet authorities seem willing in principle to sponsor and participate in such undertakings. We have alerted American economists to this opportunity through notices in the *American Economic Review* and the *Journal of Comparative Economics*, and have offered our services in establishing the necessary contacts. Thus far only two inquiries have been received, but we hope that others will follow.

LLOYD G. REYNOLDS, *Chair*

Report of the Representative to the National Bureau of Economic Research

During 1979 the National Bureau of Economic Research initiated several new large-scale research projects. Benjamin Friedman is directing a project on the changing role of debt and equity finance in the United States. A major project on inflation is being directed by Robert Hall. John Shoven, assisted by Zvi Bodie and Kim Clark, is directing a project analyzing private and public pensions, which will cut across such diverse areas of Bureau research as taxation, financial markets and monetary economics, social insurance, labor economics, and economic fluctuations. The Bureau has undertaken a simulation of the effects of changes in tax policy, allowing for behavioral response of households to changes in taxation. The Bureau is cooperating with others in Germany, Sweden, and the United Kingdom to produce comparable figures on capital taxation in the four countries. Richard Freeman is directing an NBER survey of black youths regarding causes and consequences of youth joblessness, criminal activities, and related questions.

At the July, 1979 NBER Summer Institute roughly forty faculty-level researchers worked on economic fluctuations and macroeconomic aspects of international finance. During August the Bureau's taxation project staff met to study business taxation and finance, and to begin the tax simulation and international comparisons projects described above. Next summer, month-long workshops are planned about financial markets and monetary economics, taxation, labor economics, and productivity and technical change. The Bureau's pension project staff will meet with those workshops. Shorter workshops will meet on international economics.

Bureau conferences (and organizers) in the United States and abroad in 1979 included "National Income and Product Accounts of the United States" (Murray Foss); "Youth Joblessness and Employment" (Richard B. Freeman and David A. Wise); "Economics of Information and Uncertainty" (John McCall); "Econometric Studies in Public Finance" (A. B. Atkinson, Mervyn King, and David Bradford); "International Seminar on Macroeconomics" (Robert Gordon and George de Menil); and "Taxation of Capital" (Peter Mieszkowski). The papers from all these conferences are being made available in the recently inaugurated NBER conference Paper series. It is expected that the papers resulting from the first five will be published.

In 1979 *Short-Term Macroeconomic Policy in Latin America,* edited by Jere Behrman and James A. Hanson, was published. In 1980 Ballinger will publish *Business Cycles, Inflation and Forecasting,* a collection of essays by Geoffrey H. Moore. The NBER's new primary publisher, the University of Chicago Press, will publish these books in 1980: Richard Easterlin, ed., *Population and Economic Change in Developing Countries*; Stanley Fischer, ed., *Rational Expectations and Economic Policy*; John W. Kendrick and Beatrice N. Vaccara, eds., *New Developments in Productivity Measurement*; Dan Usher, ed., *The Measurement of Capital*; and James D. Smith, ed., *Modeling the Distribution and Intergenerational Transmission of Wealth.*

Morton Ehrlich, Albert G. Matamoros, Michael Moskow, Robert R. Sterling, Don Wasserman, and Burton A. Weisbrod were elected to serve three-year terms as NBER directors.

Further information is available from Charles McLure, NBER Vice President, 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138, or the undersigned at Johns Hopkins University.

CARL F. CHRIST, *Representative*

Report of the Committee on Economic Education

Four projects dominated the attention of the Committee in 1979: the Teacher Training Program (*TTP*), the revision of the Test of Understanding College Economics (*TUCE*), the study of the economics major, and the *Journal of Economic Education*.

A workshop to disseminate the materials and instructional format of the Teacher Training Program for economics graduate students was held in the summer of 1979 at Indiana University. (See the May 1979 *Proceedings*, p. 426, for a detailed description of the *TTP*.) Phillip Saunders (Indiana University), workshop leader, was assisted by Elizabeth Alison (Harvard), William Becker (University of Minnesota), Rendigs Fels (Vanderbilt), W. Lee Hansen (University of Wisconsin-Madison), Michael Salemi (University of North Carolina-Chapel Hill), John Siegfried (Vanderbilt), John Soper (Joint Council on Economic Education), and Arthur Welsh (Joint Council on Economic Education). The workshop was oversubscribed with twenty-nine graduate students and thirty-nine faculty members attending from some twenty-six different colleges and universities. A detailed evaluation of the workshop revealed that it more than met its stated objectives. A second *TTP* workshop, under the leadership of Lee Hansen, is planned for the summer of 1980 at the University of Wisconsin-Madison. Both workshops are funded by the Lilly Endowment. With the completion of these two workshops, and the final revisions of the *Resource Manual for Teacher Training Programs in Economics*, it is expected that a substantial number of *TTP*s will be institutionalized at Ph.D.-granting institutions in the United States.

Work is continuing on the revision of the Test of Understanding College Economics. This test constitutes the primary instrument used to measure achievement and learning in economics for most of the economic-education research projects at the college and university level. The test revision, supported by a grant from the Exxon Education Foundation, is being carried out by Phillip Saunders, Rendigs Fels, and Arthur Welsh. Six forms of the test are now being normed at some thirty participating institutions. The results of this norming and a description of the revised *TUCE* will be presented at the Denver 1980 meetings of the AEA.

A new project to study the economics major was launched in 1979 by John Siegfried. The objects of the project are to a) describe the nature of the major nationally, and b) assess the agreement on the content of the major (as between faculty and students). The Alfred P. Sloan Foundation has provided a grant to support the research. Information about the major will be obtained from two sources: a departmental questionnaire, and a student questionnaire of graduating senior economics majors. Data processing will take place in the spring and summer of 1981, and a preliminary report is expected for the 1981 AEA meetings.

The Committee reviewed some of the accomplishments of the *Journal of Economic Education*, a publication established by the Joint Council on Economic Education some ten years ago. To date the *Journal* has published 21 issues with 166 different contributing authors. The circulation of 2,000 is typical of a specialty journal of this type. The *Journal* has stimulated economic-education research and has encouraged many individuals to identify with and participate in the economic-education programming of the AEA. The article, "Research on Teaching College Economics: A Survey," by Siegfried and Fels (*JEL*, September 1979) concludes that "...a cumulative literature on economic education has now developed.... [This literature] is rapidly becoming respectable, and the research findings can be useful to college economics teaching" (p. 959). Around 40 percent of the accumulated literature in economic-education research has now been published in the *Journal of Economic Education*. The continuing editorship of Henry Villard is gratefully acknowledged.

ALLEN C. KELLEY, *Chair*

Report of the Committee on Federal Funding of Economic Research

This Committee was formed by then-President Tjalling Koopmans in March 1978 as a successor to a committee which reported to the Executive Committee in 1976 (see February 1977 *Proceedings*, pp. 469–70). President Koopmans charged us to "...examine and report on federal support for economic research during the past decade, activities undertaken by other professional societies in support of research funding, and actions that the AEA should consider germane to the encouragement and support of economic education and research."

According to its certificate of incorporation, "the particular business and objects of the society" (the AEA) include: "(1) The encouragement of economic research, especially the historical and statistical study of the actual conditions of industrial life; (2) The issue of publications on economic subjects; (3) The encouragement of perfect freedom of economic discussion."

The AEA does, of course, encourage economic research by means of its meetings and publication of proceedings, which feature current research, and by sponsorship of journals and other publications. It recognizes individual researchers who have made outstanding contributions to the field of economics. Moreover, it communicates information concerning job opportunities, including research opportunities, for economists. It has an official committee to advise the U.S. Bureau of the Census, and it joins with eight other associations in forming the Committee of Professional Associations on Federal Statistics.

It names a representative to serve on the board of directors of the National Bureau of Economic Research. It also is affiliated with the Social Science Research Council and the American Association for the Advancement of Science (AAAS). It sponsors the exchange of information and ideas with economists in other nations. It recently appropriated funds to aid in seeking equal opportunities for women in the economics profession. It has sponsored special training programs for foreign and minority students, and it cooperates with the Joint Committee on Economic Education in researching and developing new educational programs in economics.

But should the AEA be asking the federal government to spend more money on economic research, or to redirect its spending on such research? Should the AEA help its members to obtain more federal funding to pursue their individual research interests? Should the AEA serve as a trade association or guild in expanding the market for its members' professional skills? Should it have a Washington office, and should it direct its officers and staff to represent the Association at various stages of the federal research-funding process?

In approaching these questions, it is helpful to have some background information concerning the present role of the federal government in funding scientific research and development (*R&D*), including that in economics.

In 1979, the federal government provided half the national total of $52 billion spent on *R&D* (NSF 78-313). The sources of funds and the location of performers of all types of *R&D* were as follows (shown in percent):

	Sources	Performers
Federal	50	13
Industry	47	71
Universities and Colleges	2	13
Other nonprofit organizations	1	3

Two-thirds of *R&D* spending was for development, 23 percent was for applied research, and 11 percent was for basic research. While the federal government provided less than half the funds for development and applied research, it funded 69 percent of the basic research. Within the federal government, the biggest spenders for

TABLE 1—ALL ECONOMISTS BY EMPLOYER TYPE, 1972

	Economists	Average Percentage of Time in *R&D*	Economists Receiving Federal Government Support
All Employees	21,120	29	8,990
Colleges or Universities	6,093	25	1,892
Government			
All	7,248		5,482
Federal		42	
State		23	
Local		30	
Nonprofit Organizations	1,882	47	969
Industry	5,439	19	667
Not Reported	458		

development were the Departments of Defense, Energy, and HEW, and the independent agency of NASA. With regard to applied research, the leaders were the Departments of Defense, HEW, Energy, and Agriculture, and the agencies of NASA, EPA, and the Nuclear Regulatory Commission. In basic research, the biggest outlays were made by the Departments of HEW, Energy, Defense and Agriculture, and the agencies of NSF and NASA (NSF 78-312).

Unfortunately, we do not have information on how the grand total of $52 billion in *R&D* funds was divided by field of science. However, we have a clue about the relative importance of economics in the total from data on full-time equivalent employees in *R&D* work. It appears that about 7,000 (or a little over 1 percent) of the total of 610,000 full-time equivalent employees in *R&D* work in 1976 were economists. This unofficial guess for 1976 is based on a 1972 estimate that the "scientist and engineer population" in 1972 included 21,738 economists. (Of these, 7,626 had a Ph.D. degree and 7,630 had a Masters degree. The AEA had 17,286 members in 1972.) Economists as a group spent 29 percent of their time on *R&D* work and another 17 percent of their time on "statistical work." Forty-three percent of economists received federal government support in 1972 (NSF 75–313 and NSF 75–327). See Table 1.

Putting development aside, total research outlays amounted to 34 percent of $52 billion or $18 billion in 1979. Of this amount, $10.5 billion was provided by the federal government (NSF 78–313). Most of the federal funds were actually spent by research performers located outside the federal government (NSF 78–312). See Table 2.

We do not have information on how the total of $18 billion was divided by field of science. However, we do know that of the $10.5 billion provided for research by the federal government, only half a billion dollars was for social science research,[1] and of that amount $176 million was devoted to economics. See Table 3. Economists may, of course, be employed on what is classified as

[1]This figure of $500 million for social science research may be compared to almost $1 billion for "social and behavioral research," or alternatively, about $1.5 billion for "social *R* and *D*". The Stokes Study Project found that, in 1976, twenty federal agencies spent $1,215 million on "social knowledge production" and $598 million on "social knowledge application." They defined knowledge production to include basic, applied, and policy research, demonstrations for policy formulation, program evaluation, and general purpose statistics. And they defined knowledge application to include demonstrations for policy implementation, development of materials, and dissemination of information. (*The Federal Investment in Knowledge of Social Problems*, Washington, National Academy of Sciences 1978, Table 1, p. 18.)

TABLE 2—FEDERAL OBLIGATIONS FOR RESEARCH BY PERFORMER, 1979

	Thousands of dollars
Intramural	3,413,434
Industrial	2,062,774
FFRDC-Independent	255,338
Universities and Colleges	3,218,858
FFRDC-Universities	696,152
Other nonprofit	599,963
FFRDC-Other	53,989
State and local governments	99,137
Foreign	83,022
Total	10,482,667

Note: FFRDC means federally funded *R&D* centers.

TABLE 3—FEDERAL FUNDS FOR RESEARCH BY SELECTED FIELDS OF SCIENCE, 1977[a]

	Total Research	Basic	Applied
All Fields	10,482,667	3,636,729	6,845,938
Social Science	556,908	128,491	428,417
Economics	176,258[b]	34,346	141,912

[a]Shown in thousands of dollars.
[b]Of this amount, $42 million was spent at universities and colleges.

other than economic research, and non-economists may be employed on economic research. Eighty-one percent of the latter amount was for what the departments and agencies classified as applied, as opposed to basic, research (NSF 78–312).

The great bulk of the federal funds for applied research in economics was provided by the Departments of Agriculture, Labor, HUD, HEW, Commerce, and State, in that order. Most of the basic research funds came from the Department of Agriculture, NSF, the Departments of HEW and Commerce, in that order (NSF 70-312, 72-312, 74-312, 76-312, 77-312, and 78-312). (See Table 4.)

Over the period 1969–79 federal funds for economic research have a little more than doubled in current dollars. (This time-series excludes all funds for development and all nonfederal funds for research.) In that same period, the *GNP* deflator went from 87 to 165. Hence, it appears that real resources devoted to such federally funded research have remained roughly constant. However, real resources declined for basic research at the same time that they rose for applied research. The same pattern is found for social science research and research in all fields of science combined as well as in economics. (See Table 5.)

The trend within economics is the result of changes in funds provided by the several leading departments and agencies (NSF 70-312 and 78-312). With respect to total funds for economic research, HUD and State increased faster than the average, while HEW and Commerce lagged. With respect to basic economic research, Agriculture and NSF led the way, while Labor, HEW, and Commerce were the slow ones. With respect to applied research, Labor, Commerce, HUD, State, and NSF increased the most, while Agriculture fell behind.

A good many public policy issues arise, of course, concerning the role of *R&D* in the national economy. The *R&D* spending was 2.2 percent of *GNP* in 1979, well below its high of 3.0 percent of *GNP* in 1964 (NSF 78-313). What is the appropriate level for the 1980's?

Federal spending was 50 percent of total *R&D* spending in 1979, while it was 70 percent of such spending in 1964 (NSF 78-313). What should this ratio be in the future?

What is an appropriate mix of spending with respect to development, applied research, and basic research? How should spending be distributed among and within the various fields of science?

How should the federal government distribute its funds among in-house researchers, those in private industry, those in universities and other nonprofit organization? To what degree should federal research funds be managed by mission agencies and to what degree by research organizations like NSF? To what extent should research topics be relevant to the needs of federal agencies as opposed to the advance of scientific disciplines?

How can "research planning" by oversight organizations and intermediation by "knowledge brokers" help to bridge gaps between policymakers and researchers?

TABLE 4—FEDERAL OBLIGATIONS FOR ECONOMIC RESEARCH BY DEPARTMENT AND AGENCY, 1979
(Thousands of dollars)

	Total Research	Basic	Applied
Departments			
Agriculture	65,881	15,134	50,747
Commerce	11,544	2,200	9,344
Defense	1,142	51	1,091
Energy	1,330	–	1,330
HEW	20,519	4,826	15,693
HUD	16,543	–	16,543
Interior	1,123	180	943
Justice	228	–	228
Labor	22,572	426	22,096
State	8,315	–	8,315
Transport	1,290	–	1,290
Agencies			
ACIR	832	–	832
App. Reg. Com.	600	–	600
CAB	836	–	836
EPA	2,700	–	2,700
FCC	1,796	–	1,796
FHLBB	1,316	–	1,316
FTC	1,319	444	875
ICC	35	–	35
NASA	520	88	432
NSF	12,642	10,947	1,695
SBA	55	–	55
Arms Control	454	–	454
International Trade Commission	2,666	–	2,666
Total, All Departments and Agencies	176,258	34,346	141,912

TABLE 5—FEDERAL FUNDS FOR RESEARCH, BASIC AND APPLIED, BY SELECTED FIELDS OF SCIENCE, SELECTED YEARS[a]

	1969	1971	1973	1975	1977	1979	1979 ÷ 1969
All Fields of Science							
Total	5,236	5,548	6,115	7,114	8,687	10,483	2.0
Basic	2,425	2,434	2,640	3,087	3,756	3,637	1.5
Applied	2,811	3,114	3,474	4,027	4,931	6,846	2.4
Social Sciences							
Total	220	306	296	303	443	557	2.5
Basic	72	70	78	73	95	129	1.8
Applied	148	236	218	231	349	428	2.8
Economics							
Total	78	75	88	127	143	176	2.3
Basic	26	25	22	22	29	34	1.3
Applied	53	50	67	105	114	142	2.7

[a]Shown in millions of dollars.

TABLE 6—FEDERAL FUNDS FOR ECONOMIC RESEARCH, BASIC AND APPLIED BY SELECTED DEPARTMENTS AND AGENCIES, 1969 AND 1979
(Millions of dollars)

	1969	1979	1979 ÷ 1969
All Departments and Agencies, Total	78	176	2.3
Basic	26	34	1.3
Applied	53	142	2.7
Department of Agriculture, Total	30	66	2.2
Basic	9	15	1.7
Applied	21	51	2.4
Department of Labor, Total	9	23	2.5
Basic	2	1	.5
Applied	7	22	3.1
Department of HEW and OEO, Total[a]	12	21	1.8
Basic	6	5	.8
Applied	6	16	2.7
Department of Commerce, Total	6	12	2.0
Basic	3	2	.7
Applied	3	9	3.0
Department of HUD, Total	2	17	8.5
Basic	–	–	–
Applied	2	17	8.5
Department of State, Total	2	8	4.0
Basic	–	–	–
Applied	2	8	4.0
NSF, Total[b]	5	13	2.5
Basic	5	11	2.4
Applied	1	2	3.4

[a]The OEO total was 5 in 1969 and zero in 1979.
[b]The ratios for NSF are based on unrounded data.

How should a government agency select a research contractor? What degree of management control and monitoring of the research performance of grantees and contractors is necessary and productive? How can a government agency evaluate the quantity and quality of the research output it has purchased?

Is it important to distribute federal funds for *R&D*, and the capacity to perform *R&D*, more evenly than it now is by region and state?

Economists are perhaps as well-equipped as others to participate in the answering of those public policy questions. Some of them, indeed, are specialists in the study of those questions. Others sit in positions with authority to decide how to allocate taxpayer money or company resources for the funding of economic research. Many members of the economics profession are both producers and consumers of economic research. They are both suppliers of research time and effort, and demanders of it.

Even though members of the AEA are closely involved in the several aspects of research funding, oganization, utilization, and evaluation, it does not necessarily follow that the Association should adopt an organizational position on all or any of the policy questions listed above. While the AEA is dedicated to the encouragement of

economic research, it has declared that it "will take no partisan attitude, nor will it commit its members to any position on practical economic questions."

As we noted at the outset of this report, the AEA does encourage research in a number of ways, but it is, relative to some professional and interdisciplinary and research-performer organizations, quite uninvolved with attempts to influence the federal funding and direction of economic research. Some professional organizations do direct lobbying aimed at increasing federal funds for research by members of their organizations. Some also affiliate themselves with umbrella organizations which serve as pressure groups on behalf of more federal and other research funding.

We have reviewed several possible actions that the AEA might consider, to use President Koopman's words, "germane to the encouragement and support of economic education and research." Three such actions which we have reviewed but do *not* recommend are the following:

1) The AEA could seek a stronger role as a constituent organization in AAAS and in other interdisciplinary societies which lobby for research and development funds. This might involve the Association in taking a partisan attitude on issues far removed from economic research.

2) The AEA could produce and publish a monthly newsletter to communicate information to the membership about the "Washington jungle" of research funders, intermediaries, performers, and users. It might be predominantly a staff effort modeled after the newsletter published under the title of *Footnotes* by the American Sociological Association. We are not aware of a critical need for such a publication.

3) The AEA could establish a committee to study (a) the research activities of and funding sources used by economists, with attention to their subfields of interest, type of employment, age and other personal traits; (b) the methods by which economists gain information about and seek research funds; and (c) the uses of research outputs by government agencies and others. We see this as an interesting project, but one with limited payoff.

We offer the following recommendations.

(1) *That the AEA offer to appoint committees to advise federal departments and agencies with respect to their primary data collection and presentation of basic economic indicators.* The Association presently does appoint members to the Committee of Professional Associations on Federal Statistics and the AEA Advisory Committee to the Bureau of the Census. We believe that the Bureau of Labor Statistics and several other federal agencies might welcome a similar association with AEA and that this would be a useful activity for the Association.

(2) *That the AEA appoint a standing committee to serve as a line of communication with the Economics Program of the Social Science Division of NSF.* We believe that the basic research funded by this program is of unique importance to the overlapping interests of the AEA and the federal government in advancing the discipline of economics. At the same time, this type of research is most likely to be underfunded. We believe that such an AEA committee, by supplying members of the Association with information about this NSF Program, could assist them in their individual efforts to defend and improve that program.

ROBERT LAMPMAN, *Chair*

Report of the Census Advisory Committee

The Committee Report consists primarily of a set of recommendations made to the Director of the Census at our last meeting, to wit:

> The Committee strongly urges the Census Bureau to give very high priority in its current and future planning to increasing our understanding of the nature of the measured employment and unemployment of teenagers, minorities, women, and other groups with special employment problems of particular importance to economic policymakers. In addition, we believe that a very high priority be given (a) to the development of improved 'gross flow' data (per the Gordon Commission Report and the Report of the National Commission on Employment and Unemployment Statistics), and (b) to the improvement of our understanding of seasonal factors in employment. Among the vehicles that may be used to make some progress on these issues is the use of additional and redesigned questions in the *Current Population Survey*. Such new questions could be included in the Special Sample of about 10,000 households recommended by the NCEUS.
>
> The Committee requests that at the next meeting of this Committee, the Census report on the state of 'gross flow data' as well as on the problems that may arise in implementing the recommendations above.

The Committee is concerned about response biases and errors in gross flow data, particularly youth unemployment data bias, which appears to be more than just a sample design problem. The Committee felt these are matters of considerable importance to national economic and social policy and suggests that the Executive Committee might wish to consider some action on its part, possibly in the form of a resolution to be submitted to the membership and published in the *Proceedings*.

The Bureau of the Census is inviting suggestions from data users regarding the mid-decade census, with a deadline of June 1980, which might be noticed in some form in the *American Economic Review* by the Executive Committee. In this connection, the Committee has some concerns about the mid-decade census plans, for example, the purposes and justification for that census, interpretation of the "enumeration" requirement, the possibility of utilizing sampling procedures, and the importance of recognizing tradeoffs between accuracy of counts and economy. Our Committee will continue to explore these questions with the Bureau, but at the same time feels the Executive Committee should be aware of our concerns and take whatever action it considers appropriate.

The Committee also would like to advise the Executive Committee of the ongoing American Statistical Association/National Science Foundation Census Bureau Fellowship Program for faculty with special interest and background in the area of economic data and methodology. The fellowships pay up to $50,000 per year to be spent at the Bureau. They also provide some research assistance support. Further information may be obtained from Fred Leone, Executive Director, American Statistical Association, 806 15th Street, N.W., Washington, D.C. 20005; telephone (202)393-3253.

Finally, the Committee suggests that the Executive Committee explore the possibility of appointing an Advisory Committee to the Bureau of Labor Statistics with similar functions to that of the Census Advisory Committee. In our opinion, such a committee would be useful in connection with some of the issues mentioned above, as well as related Bureau of Labor Statistics data collection and survey activities. We believe Secretary Ray Marshall would likely respond favorably if approached by the Association to establish an AEA/BLS advisory committee along the lines suggested.

ROBERT F. LANZILLOTTI, *Chair*

Begin making plans to attend the

Ninety-Third
Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

DENVER, COLORADO

September 5-7, 1980

The 1980 Professional Placement Service will be held December 28-30 at the Dallas Hilton, Dallas, Texas. Please note that there will not be a Professional Placement Service in Denver at the September meetings.

See the Notes section of the June *AER* for the American Economic Association's preliminary program.

The 1981 meeting will be held in Washington, DC, December 28-30.

---

2 NEW JOURNALS FROM NORTH-HOLLAND PUBLISHING COMPANY

CARNEGIE-ROCHESTER CONFERENCE SERIES ON PUBLIC POLICY

Editors: KARL BRUNNER, University of Rochester, and ALLAN MELTZER, Carnegie-Mellon University

The Center for Research in Government Policy and Business at the U of R and the Center for the Study of Public Policy at C-MU have jointly, since 1973, developed a series of semi-annual conferences explicitly designed to direct attention to major issues in economic policy. They are used to survey in detail our knowledge pertaining to selected policy problems or to explore the nature and weight of important issues confronting policy makers, and are based on research papers specifically prepared by qualified scholars. Each conference concentrates on a unified subject or problem group; the separate volumes thus tend to cover a coherent set of problems. For the past four years, the volumes have been distributed as a supplementary series to the Journal of Monetary Economics. As from 1980, the **Carnegie-Rochester Conference Series** will be published as a journal in its own right.

1980: Volumes 12 and 13 (600 pages). Subscription price: US $86.75/Dfl. 178.00 including postage.

JOURNAL OF ECONOMIC BEHAVIOR AND ORGANIZATION

Editors: RICHARD H. DAY, University of Southern California, and SIDNEY G. WINTER, Yale School of Organization and Management

The **Journal of Economic Behavior and Organization** is devoted to theoretical and empirical research concerning economic decision, organization, and behavior. Its specific purpose is to foster an improved understanding of how human cognitive, computational, and informational characteristics influence the working of economic organizations and market economies. The JEBO is eclectic as to research method; systematic observation and careful description, experimental and ethological study, simulation modeling and mathematical analysis are all within its purview. Empirical work that probes close to the core of the issues in theoretical dispute is encouraged.

1980: Volume 1 in 4 issues (400 pages). Subscription price: US $66.25/Dfl. 136.00 including postage.
Private subscribers are entitled to a subscription at the reduced rate of US $37.00/Dfl. 76.00 including postage.

Personal subscriptions must be prepaid and sent directly to the Publisher. Copies should not be made available to institutions.

The Dutch guilder price is definitive. US $ prices are subject to exchange rate fluctuations.

Orders and requests for specimen copies may be sent to the Publisher:

NORTH-HOLLAND PUBLISHING COMPANY

P.O. Box 211 - 1000 AE Amsterdam - The Netherlands
52 Vanderbilt Avenue - New York, N.Y. 10017 - U.S.A.

0710 NH

CAMBRIDGE

Inflation and the Personal Income Tax
An International Perspective

Vito Tanzi. A detailed analysis of the impact of inflation on income tax systems in many countries. Tanzi examines some of the more important distortions of income tax systems caused by inflation and discusses possible corrective measures, comparing them with the actual experiences of the United States, United Kingdom, France, Australia, Canada and others in their pursuit of policies aimed at neutralizing the impact of inflation. $27.50

The Multinational Corporation
A Radical Approach

Stephen Hymer. Edited by Robert B. Cohen, Noel Dennis, Nadine Felton, Morley Nkosi, and Jaap van Liere. This book brings together Hymer's major writings on the multinational corporation. An introductory essay explains Hymer's relationship to both mainstream and Marxian analysis of the multinational corporation, and a brief biography and complete bibliography conclude the book. $35.00

Profitability and Unemployment

Edmond Malinvaud. An examination of the rationale for the current argument in economic policy which states that an adequate level of profitability is a precondition for avoidance of future unemployment. Malinvaud's analysis is built around a simple model of the Harrod-Domar type, and extends the study of fixed price equilibria through a consideration of a dynamic process whose major components are capital accumulation and shifts in income distribution. A co-publication of *Cambridge University Press* and *Les Editions de la Maison des Sciences de l'Homme.* $12.95

The New International Division of Labour

Folker Fröbel, Jürgen Heinrichs, and Otto Kreye. Translated by Pete Burgess. The authors contend that, because of the changing world economic structure, companies have been forced to reorganize production on a global scale, and argue that prevailing conditions in the capitalist world economy mean that individual nations' efforts to control industrialization and unemployment are doomed to failure. *Studies in Modern Capitalism/Etudes sur le Capitalisme Moderne.* A co-publication of *Cambridge University Press* and *Les Editions de la Maison des Sciences de l'Homme* $54.50

The Economics of Individual and Population Aging

Robert L. Clark and Joseph J. Spengler. A study focusing both on the individual and on the macroeconomic problems caused by the increasing proportion of the aged in the population. *Forthcoming*

The Collected Writings of John Maynard Keynes

Donald Moggridge, Editor

Volume 24: Activities 1944–46: The Transition to Peace $35.00

Volume 29: The General Theory and After: A Supplement $35.00

Essays on John Maynard Keynes

Milo Keynes, Editor *Now in Paperback* $9.95

Economics and Consumer Behavior

Angus Deaton and John Muellbauer

The authors have used duality to provide an integrated and accessible treatment of the economics of consumer behavior. This book focuses on applications of the theory to welfare economics and econometric analysis. Much of the material is too recent to have appeared in any previous text. The structure of the book is designed to permit its use in a wide variety of courses at different levels and examples are provided throughout. Hardcover $49.50 Paper $18.95

Contents

Part One:
Consumer Demand Analysis

- The limits to choice
- Preferences and demand
- The theory at work
- Extensions to the basic model

Part Two:
Separability and Aggregation

- Restrictions on preferences
- The theory of market demand

Part Three:
Welfare and Consumer Behavior

- Consumer index numbers
- Household characteristics, demand, and household welfare comparisons
- Social welfare and inequality

Part Four:
Extensions and Applications

- The quality of goods and household production theory
- Labor supply
- The consumption function and intertemporal choice
- The demand for durable goods
- Choice under uncertainty

Theory of Value, Prices and Accumulation

A Mathematical Integration of Marx, von Neumann and Sraffa

Gilbert Abraham-Frois and Edmund Berrebi

Translated by M. P. Kregal-Javaux. A discussion of the body of economic theory that has been built upon foundations laid by Sraffa. The material is presented in a synthetic manner and is related to the work of von Neumann and Marx, resulting in a general overview of the modern work on value, prices, and growth which has developed from these three major precursors. $24.95

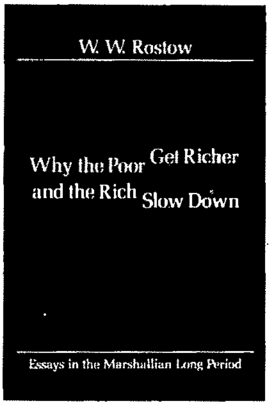
Revealed Preference of Government

Kaushik Basu

A critical examination of standard methods for discovering governmental preferences which proposes a generalized method that can accommodate a much larger range of behavior. Basu draws upon a variety of theoretical tools from welfare economics and other areas of the social sciences. $22.50

Cambridge University Press

32 East 57th Street, New York, N.Y. 10022

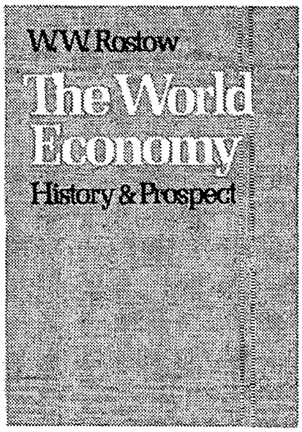


Why the Poor Get Richer and the Rich Slow Down

ESSAYS IN THE MARSHALLIAN LONG PERIOD

By W. W. Rostow

Bearing on fundamental issues of economic theory, history, and public policy, this volume elaborates and goes beyond themes enunciated in W. W. Rostow's previous works. The eight essays presented here are unified by the author's insistence that neo-Keynesian and neoclassical theory are an inadequate basis for economic analysis and policy prescription. 456 pages; $19.95



The World Economy

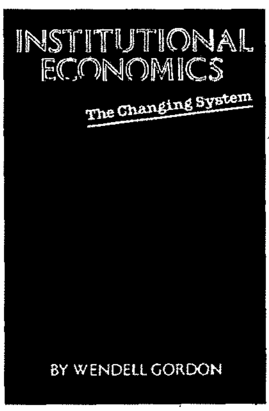
HISTORY AND PROSPECT

By W. W. Rostow

"The scope of the book is astonishing. . . . it traces the course of the world economy since the 18th century."
—Lindley H. Clark, Jr., *Wall Street Journal*

". . . the path of evolution from the pre-industrial to the advanced industrial condition is meticulously chronicled."
—Jock Bruce-Gardyne, *Daily Telegraph*

1978, 877 pages, 275 tables, 157 charts; $34.50



Institutional Economics

THE CHANGING SYSTEM

By Wendell Gordon

This volume clearly and evocatively sets out the philosophy of economic institutionalism, which, in Wendell Gordon's view, offers the most hope for understanding and solving the economic problems of the twentieth century. Gordon's argument that the dynamics of technical change, the institutionalism of behavior norms, human biology, and the resource endowment all interact in an ongoing social process is a central tenet of the institutional approach.

400 pages, 15 figures, 4 tables; $9.95, paper; $19.95, cloth

Toward a Reasonable Society

By C. E. Ayres

"Professor Ayres provides in this book a rigorous account of our world. . . . By appealing wholeheartedly to the scientific-technological bases of life, [he] develops a sanguine philosophy, liberal and liberating."
—*Technology and Culture*

1961, 309 pages; $7.95, paper

Science and Ceremony

THE INSTITUTIONAL ECONOMICS OF C. E. AYRES

Edited by William Breit and William Patton Culbertson, Jr.

". . . *Science & Ceremony* has reawakened interest in one of the most striking characters in the economics profession of the United States in this century."
— Kenneth E. Boulding

1976, 228 pp.; $14.95

071

TEXAS

To order directly send payment to: **University of Texas Press**
Texas residents add 5% sales tax. POST OFFICE BOX 7819 AUSTIN, TEXAS 78712

Economic Issues of the Eighties

edited by Nake M. Kamrany and Richard H. Day

This volume gathers the insights and talents of a number of leading scholars and economists in discussions of such issues as the "stagflation" trap, the spiraling costs of government-financed health care, the downturn in productivity, the fluctuations of the money market, and other economic issues of the 1980s. $6.95 paperback, $20.00 hardcover

Productivity in the United States

Trends and Cycles

John W. Kendrick and Elliot S. Grossman

In this timely study, John Kendrick and Elliot Grossman examine the course of productivity changes in the U.S. economy since 1948 and clearly explain the relationship between fluctuations in productivity and inflation. *Productivity in the United States* can serve as a handbook for those with a scholarly or professional interest either in changes in the American economy as a whole or in a particular industrial sector. $14.95

State Income Taxation

Clara Penniman

Clara Penniman describes the present-day state income tax administration and defines the features that make these taxes the main producers of state revenues. Comprehensive in scope, the book covers statutory policies, administrative organization, management resources, compliance, collections, and the auditing of personal and corporate returns. $22.50

New in paperback

The Theory of Money

Jürg Niehans

"*The Theory of Money* is a comprehensive and elegant elaboration of the neoclassical tradition in monetary theory, synthesizing and extending Niehans's earlier work in this field....Niehans's presentation is highly structured, invariably lucid, and rigorously compact....His careful distangling of traditional debates, augmented by an extensive bibliography, suggest its usefulness as an advanced textbook."—*Journal of Economic Literature*

Exploring the function of money as a medium of exchange and a store of value, this major work provides a comprehensive introduction to the pure theory of money and monetary policy. $6.95 paperback, $17.95 hardcover

Johns Hopkins

The Johns Hopkins University Press
Baltimore, Maryland 21218

New from Martinus Nijhoff Publishing

Limits to the Welfare State

Gerard Jan van Driel, Jacob A. Hartog, and Cornelius van Ravenzwaaij
Erasmus University, Rotterdam
The conflicting demands posed by a modern industrial society on the resources of its economy are studied in this book by means of a dynamic 22-sector (Leontief-type) model.

1980, cloth, 192 pp.
$19.50/Dfl. 42.90

Welfare Economics and Social Choice Theory

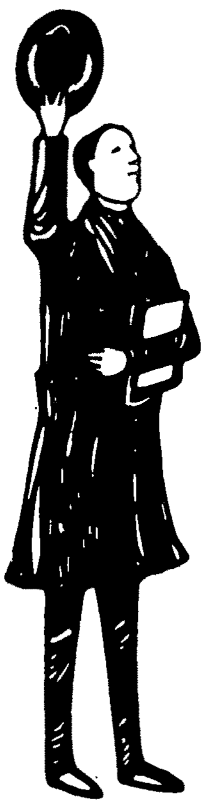
Allan M. Feldman
Brown University
"Is a particular economic or voting mechanism good or bad for society?" Feldman's book examines this question by exploring the major topics of welfare economics and social choice theory.

1980, cloth, 256 pp.
$26.00/Dfl. 60.00
(paper $12.00)

World Communism at the Crossroads
Communist Military Ascendancy, Political Economy, and Human Welfare

Steven S. Rosefielde, editor
University of North Carolina, Chapel Hill
This collection of essays documents the newly attained military ascendancy of the Communist bloc, and attempts both to analyze the aspects of communist political economy which have made this achievement possible, and to assess how the issue of human welfare fits into the tapestry of communist development.

1980, cloth, 336 pp.
$19.95/Dfl. 45.00



Energy Policy Modeling
United States and Canadian Experiences, Volumes 1 and 2

edited by
W. T. Ziemba and S. L. Schwartz
University of British Columbia
E. Koenigsberg
University of California, Berkeley
An extensive report from the Canadian Energy Policy Modeling Conference on how energy modeling can make a contribution to the energy policy decision-making process. The two volumes explore both specialized energy policy models, and attempts to mesh supply and demand elements in an integrative fashion.

Volume 1:
1980, cloth, 400 pp.
$35.00/Dfl. 80.00
Volume 2:
1980, cloth, 368 pp.
$35.00/Dfl. 80.00

To order, or to request a free catalog, please write to:

Martinus Nijhoff Publishing
Suite 44D
160 Old Derby St.
Hingham, MA 02043

Orders from outside North America should be sent to:

Martinus Nijhoff Publishers BV
Lange Voorhout 9-11
P.O. Box 566
2501 CN The Hague
The Netherlands

Individual orders must be prepaid.
Prices are subject to change.

Economic Policy and the Great Stagflation

By ALAN S. BLINDER

Economic Policy and the Great Stagflation presents a historical analysis of the major macroeconomic events in the United States between 1971 and 1976. The book examines the combined recession and inflation of this period, focusing on and critically treating the effects of government fiscal and monetary policies and wage–price controls.

Introductory chapters outline the elementary theory of supply shocks and stagflation and provide a historical overview of the period. The Great Recession is evaluated, with a corresponding look at the stunning acceleration and deceleration of inflation during the years from 1973 through 1975. The study pays particular attention to three "supply shocks": OPEC, rising food prices, and the end of wage–price controls. Later chapters feature new econometric evidence on the effects of the price controls programs and the 1975 tax rebate.

1979, 248 pp., $19.50 ISBN: 0-12-106160-4

Economic Analysis of an Urban Housing Market

By JOHN F. McDONALD

A Volume in the STUDIES IN URBAN ECONOMICS Series

CHAPTER HEADINGS: Introduction. Models for the Demand for Urban Housing. The Value of Commuting Time. Race Externalities, the Local Public Sector, and the Demand for Housing. Models of the Supply of Urban Housing. Analytical Models of the Urban Housing Market. Population Density in Chicago. Land Values in Chicago. A Model of the Chicago Housing Market.

References. Index.

1979, 224pp., $19.50 ISBN: 0-12-483360-8

Future volumes in the STUDIES IN URBAN ECONOMICS Series are now available on a Continuation Order basis. Your Continuation Order authorizes us to ship and bill each future volume in the series automatically, immediatedy upon publication. This order will remain in effect until cancelled. Specify the volume number or title with which your order is to begin.

Regulatory Change in an Atmosphere of Crisis

Current Implications of the Roosevelt Years

Edited by GARY M. WALTON

CHAPTER HEADINGS: *G. M. Walton*, Monetary Crises and Government Reliance in Modern Times. *J. M. Buchanan*, The Economic Constitution and the New Deal: Lessons for Late Learners. *J. Hughes*, Roots of Regulation: The New Deal. *K. G. Elzinga*, The Robinson–Patman Act: A New Deal for Small Business. *H. Jones*, Banking Reform in the 1930s. *R. J. MacKay and J. D. Reid., Jr.*, On Understanding the Birth and Evolution of the Securities and Exchange Commission: Where Are We in the Theory of Regulation? *E. K. Browning and W. R. Johnson*, Taxes, Transfers, and Income Inequality. Subject Index.

1979, 184 pp., $16.00 ISBN: 0-12-733950-7

Welfare of the Poor

By MARY BRYNA SANGER

CHAPTER HEADINGS: WELFARE CASELOAD DYNAMICS. Caseload Variations: The Conceptual Issues and Methodological Techniques. Variations within the Welfare Population. EMPLOYMENT. Measuring Employability. Factors Affecting Work and Welfare Patterns. FAMILY COMPOSITION. Introduction. Determinants of Marital Instability. The Effects of the Welfare System on Remarriage. The Effects of the Welfare System on Illegitimacy. The Effects of the Welfare System on Fertility. HEALTH. Health Definitions. Nutrition: A Causal Factor. Utilization of Health-Care Services among the Poor. EDUCATION. Introduction. Socioeconomic Determinants. Inequality of School Resources. Institutional Determinants. WELFARE RESEARCH: POLITICAL AND METHODOLOGICAL LIMITATIONS. Impact of Different Paradigms on Understanding Dependency. Cost of Knowledge. Policy Implications and Future Research on Poverty. References.

Conclusions appear at the end of each chapter.

1979, 192 pp., $15.00 ISBN: 0-12-618650-2

Send payment with order and save postage and handling charge.
Prices are subject to change without notice.

ACADEMIC PRESS, INC.

A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

NEW AND RECENT BOOKS FROM MIT

The Collected Papers of Franco Modigliani
Volume 1:
Essays in Macroeconomics
Volume 2:
The Life Cycle Hypothesis of Saving
Volume 3:
The Theory of Finance and Other Essays
edited by Andrew Abel

These three volumes include the major scientific contributions of one of the world's foremost economists. These 45 essays span nearly four decades. Each volume is equipped with its own preface, introduction by Modigliani, and name index.
$25.00 each volume

Collected Economic Papers of Joan Robinson
Five Volumes

Joan Robinson has earned an international reputation as a radical critic of neoclassical theory. Her unique contribution, according to an article in *Challange,* "lies not only in her recognition of the full significance of the Keynesian revolution, and its relationship to Marxist theory, but also in the manner in which, on these bases, she has forged new theoretical tools to tackle a remarkable variety of economic problems."

When the first volume of Robinson's papers appeared in Britain, *The Economist* remarked that "it has all the Robinson hallmarks; the zestful virtuosity in the handling of abstract ideas, the lively style, the flashes of irreverence, the erudition . . . which make this brilliant writer so interesting."
$25.00 each volume
Boxed set, with separately bound index, $125.00

Introductory texts available in paperback:

The Economic Theory of Pollution Control
by Paul Burrows

This book covers recent developments in the economic analysis of pollution without the use of mathematical models, making it accessible to a wide range of economics students.
$8.95
(cloth edition, $25.00)

A Guide to Econometrics
by Peter Kennedy

"The writing is clear and lively and the content is interesting. . . . The author is quite successful in guiding the reader through the subject without much effort, and in embellishing the process with interesting insights and useful references."—
Jan Kmenta, University of Michigan
$7.95
(cloth edition, $19.95)

The MIT Press

Massachusetts Institute of Technology
Cambridge, Massachusetts 02142

There's still time to adopt—

Roger N. Waud's
ECONOMICS

The first introductory text that reflects *today's* profound and dramatic economic changes. Given recent shocks to the economy, the author's response has been to work aggregate supply into macroeconomic analysis. In the macroeconomics area you'll find a new emphasis on aggregrate supplies...wage-price guidelines and controls...methods of dealing with unemployment...discussions of the 60's and 70's manpower programs...and much more.

In the microeconomics area, Professor Waud focuses on pure competition compared with monopoly...oligopolistic market structures...the relationship between market structure and wage determination...income distribution, poverty, and welfare policy...such applications of basic supply and demand as rent controls, agricultural price supports, and externalities...and other current issues.

Supplements include an instructor's manual with test bank, transparency masters, a study guide, and ACCESS, a unique computer-generated test system.

ISBN 0-06-056963-3. 848 pages. January, 1980.

Douglass C. North and Roger LeRoy Miller's
ECONOMICS OF PUBLIC ISSUES

Fifth Edition

Now updated and featuring eighteen new essays, this bestselling supplementary text puts your students right in the middle of today's news. New topics include advertising, drug regulation, economics of the future, international cartels, revival of the draft, TV regulation, private fire fighters, wage and price controls, illegal aliens, the social security system, black economic progress, educational vouchers, animal extinction, smoking, and the value of human life.

An instructor's manual is also available.

ISBN 0-06-044875-X 214 pages March, 1980.

Lawrence S. Zudak's
MANAGERIAL ECONOMICS

The depth of coverage, variety of examples and problems, and quality of analytical materials place this new text in a class of its own. See why one reviewer claimed, "A first-rate book. In fact it could be the best in the field."

ISBN 0-06-047451-3. 566 pages. January, 1980.

To request examination copies, write to Suite 5D (886), Harper & Row, 10 East 53d St., New York, N.Y. 10022. Please include course title, enrollment, and present text.

Harper & Row

New York • San Francisco

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

AEA sponsored Group Life Insurance for you and your family— at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA has joined a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by dividend credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 45% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future dividends and credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program B-1
1707 L Street, N.W.—Suite 700
Washington, D.C. 20036

Please send me more information about the AEA Life Insurance Plan.

Name ______________________ Age ______

Address ______________________

City ______________ State ______ Zip ______

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Please mention The American Economic Review *When Writing to Advertisers*

Wiley Books Mean Business, But Not As Usual...

PUBLIC AND PRIVATE INVESTMENTS:
Socioeconomic Analysis
Henry M. Steiner

Here's a book that describes the correct methods of judging capital investments. It's the only text that attempts to link project analysis with traditional microeconomic analysis. The first part of the book deals with decision making when the price system is assumed to be working perfectly. Part two discusses the meaning of resource allocation under the price system, particularly when the price system is not working. The third part examines the socioeconomic analysis in the development of nations.
(0 471 01625-X) 1979
approx. 450 pp. $24.95 (tent.)

MULTINATIONALS UNDER FIRE:
Lessons in the Management of Conflict
Thomas N. Gladwin and Ingo Walter

For the first time, a book that presents a comprehensive theoretical and practical analysis of multinational corporate conflict in a managerial framework. Issues ranging from human rights, terrorism, environmentalism and consumerism to labor relations, tax avoidance, and bribery/corruption in international business are discussed in detail. In each case, the issue is analyzed in cause-and-effect terms from the vantage-point of the multinational enterprise. Concrete cases anchor the concepts in real-world events. A new method of conflict diagnosis, planning and policymaking is also presented.
(0 471 01969-0) Feb. 1980
712 pp. $32.95

FORECASTING METHODS FOR MANAGEMENT, 3rd Ed.
Steven C. Wheelwright and Spyros Makridakis
(Wiley Series on Systems and Controls for Financial Management)

Although most managers are aware of the need for improved forecasting, few are familiar with the range of techniques that have been developed and the characteristics that must be known to select the most appropriate technique in a given situation. Here's a completely revised and up-dated edition that covers a wide range of forecasting methods and describes how they can be applied in practice, but does not dwell on the theoretical questions behind their development.
(0 471 05630-8) March 1980
approx. 300 pp. $24.95 (tent.)

UNPRICED VALUES:
Decisions Without Market Prices
J. A. Sinden and Albert C. Worrell

This book deals with the problem of placing comparative values on benefits and costs for which there are no market prices (e.g., scenic beauty, wildlife preservation, human health, outdoor recreation, etc.). The book, which demonstrates how decisions can be made in a rational manner, is the first comprehensive and systematic treatment of information on this subject. The methods that have been developed are organized into logical groups and they are offered in such a way that they can be useful to workers in fields other than those in which they were developed. There is an analysis of the relationship between valuation and decision making, and a discussion of just how the nature of the decision determines the kind of comparative values that are needed and the valuation costs that can be justified. Includes examples taken from the environmental and natural resource area to show how the methods can be applied.
(0 471 02742-1) 1979
511 pp. $25.95

MATHEMATICAL PROGRAMMING:
Structures and Algorithms
Jeremy F. Shapiro

One of the first books to incorporate and integrate most of the central constructs and main results of mathematical programming in one volume. It features many of the important algorithms for solving mathematical programming problems. This book also shows how to cut administration costs in welfare plans and how to speed processing of claims. It emphasizes Lagrangean duality theory and related concepts of convex analysis and explains how mathematics (including linear algebra, analysis, and group theory) is applied in the analysis and solution of optimization problems.
(0 471 77886-9) 1979
388 pp. $23.95

Available at your bookstore or write to John Storojev, Dept. 7564

WILEY-INTERSCIENCE
a division of John Wiley & Sons, Inc.
605 Third Avenue
New York, N.Y. 10016
In Canada: 22 Worcester Road, Rexdale, Ontario
Prices subject to change without notice.

092 1-7564

Foreign Enterprise in Developing Countries

Isaiah Frank

Representatives of multinational corporations have traditionally stressed their contributions to Third World nations through the role of private foreign investment. On the other hand, critics charge multinationals with a long list of sins and abuses. *Foreign Enterprise in Developing Countries* synthesizes data and information obtained from nearly 100 multinational companies and some 360 of their subsidiaries in less-developed nations in an attempt to more accurately gauge the effects of foreign enterprise on developing economies. $5.95 paperback, $12.95 hardcover

Employment and the Wage Bargain

H. M. Douty

H. M. Douty, former head of the U.S. Bureau of Labor Statistics, presents a broad review of the nature and outcome of the wage determination process, especially as it has evolved in the twentieth century. Douty provides a nontechnical summary of fluctuations in the supply and demand for labor; wage differentials for skilled and unskilled workers; comparisons of direct wages and wage supplements; and discusses the relationship between wages and prices. *Policy Studies in Employment and Welfare, Number 37* $4.95 paperback, $12.00 hardcover

Community-Based Employment Programs

Howard W. Hallman

In an attempt to improve the effectiveness of local employment and training programs, Howard Hallman evaluates the political struggle, arising from CETA funding allotments, between two types of these organizations—one community-based, the other city- or county-sponsored. Hallman makes several important recommendations concerning the development of basic criteria for these programs and outlines a plan for the fair allocation of funds. *Policy Studies in Employment and Welfare, Number 36* $4.95 paperback, $12.00 hardcover

Johns Hopkins

The Johns Hopkins University Press
Baltimore, Maryland 21218

A Revealing Close-up of the World's Newest Economic Superpower.

It has virtually no army, no natural resources and no farm land to speak of. It has controlled crime and pollution. It is a free and vigorous democracy. And only the U.S. and the U.S.S.R. are more productive. Japan, the people and society that have achieved a post-war economic miracle.

JAPAN
THE FRAGILE SUPER POWER
Updated and Revised Edition
by FRANK GIBNEY

"...a work of distinction." —Theodore H. White

"Gibney is the best informed American to write on Japan since Ambassador Edwin O. Reischauer." —John Barkham Reviews

MERIDIAN F528/$5.95

New American Library Education

NAL TIMES MIRROR

1633 Broadway
New York, N.Y. 10019

THE ECONOMICS INSTITUTE
Boulder, Colorado
offering

Intensive Preparation For Post-Graduate Programs In Economics And Business For Students From Abroad

- ☐ English
- ☐ Mathematics
- ☐ Economic Theory
- ☐ Statistics
- ☐ Accounting
- ☐ Computer Applications
- ☐ Business Organization and Management

Current program dates, tuition and fee schedule, and further information available from:

The Director
Economics Institute
Campus Box 259
University of Colorado
Boulder, Colorado 80309

Sponsored by the American Economic Association

New from Columbia

The Political Economy of Regulation

Creating, Designing, and Removing Regulatory Forms

BARRY M. MITNICK. Drawing from the fields of economics, political science, and law, the author has developed a new approach to the problem of assessing the effectiveness of government regulation of business. Providing both a conceptual analysis of regulation and an integration of material not previously available in one source, the volume is a significant contribution to the literature on policy process, organization theory, and public choice. 496 pages, $20.00

Sterling-Dollar Diplomacy in Current Perspective

The Origins and the Prospects of Our International Economic Order; New Expanded Edition

RICHARD N. GARDNER. "On one level, it is a fascinating study of the politics of international economic management in the postwar years; on another, it is a trenchant analysis of the hopes and fallacies of international economic planning. The author's new introduction for the 1980 edition brilliantly assesses the successes and failures of our postwar economic institutions." —*Arthur M. Schlesinger, Jr.* 480 pages, $25.00 cloth, $7.50 paper

Benefit Plans in Higher Education

FRANCIS P. KING AND THOMAS J. COOK. Totally up-to-date and comprehensive, *Benefit Plans in Higher Education* offers concise description, analysis, and evaluation of the principal benefit plans provided by colleges and universities in the United States. The latest volume in the now standard TIAA-CREF series on the subject, this book represents a full-scale revision of *Benefit Plans in American Colleges* (1969) and includes junior and community colleges for the first time. 400 pages, $20.00

Decisions, Politics, and Change

A Study of Norwegian Urban Budgeting

ANDREW T. COWART AND KARL ERIK BROFOSS. Explores Norwegian public budgeting from the standpoint of how individuals in public institutions make allocation decisions. *Distributed in U.S. and Canada for Universitetsforlaget.* 160 pages, $11.00 paper

To order send check or money order to Dept. JN at the address below, including $1.30 per order for postage and handling.

COLUMBIA UNIVERSITY PRESS

Address for orders:
136 South Broadway, Irvington, New York 10533

THE Publisher in Business and Economics

Standing for:

Integrity in our editorial program which selects for publication only those texts that meet high standards of pedagogy and academic excellence. Irwin books can be adopted with confidence that they represent the very best in quality of subject matter and form of presentation, written by competent and reputable authors.

Reliability in providing prompt and accurate service to faculty, bookstores, and students. Our computerized system of order fulfillment assures fast shipment of all publications upon receipt of orders — faster than most other publishers.

World-wide distribution of Irwin books which are read and studied in thousands of colleges and universities in the U.S., Canada, and abroad. Authors of Irwin books can be assured that their works will be as widely read as possible throughout the world.

Innovation in pioneering many new developments in higher education for business, including publication of casebooks, programmed learning, and new series in behavioral science for business, quantitative analysis for business, and information and decision sciences.

Noteworthy for providing the best and most complete array of supplementary teaching and learning aids to assist instructors and students. Irwin texts are also revised at timely intervals to keep them up to date with the latest trends and developments.

Look to Irwin for leadership in higher education for business. Since 1933 we have been committed to publishing exclusively in our specialized areas to meet the classroom needs of three generations of students.

RICHARD D. IRWIN, INC.

Homewood · Illinois 60430

Competition in an Open Economy

A Model Applied to Canada

Richard Caves, Michael E. Porter, A. Michael Spence with John T. Scott

This book extends the structure/conduct/performance framework of analysis to present a fully specified simultaneous equation model of an open economy—Canada. By estimating a system of equations of major variables, authors identify dependent and independent variables and assess the effect of structural variables on diversification strategies and the consequences for market performance. *Harvard Economic Studies, 150* $25.00

Essays in the Economics of Uncertainty

Jean-Jacques Laffont

These three elegant essays develop principles central to the understanding of the diverse ways in which imperfect information affects the distribution to resources, incentives, and the evaluation of economic policy. The first concerns the special role that information plays in the allocation process, the second essay deals with the extent to which agents can influence the random events they face, and the final essay discusses the problem of certainty equivalence in economic policy. *Harvard Economic Studies, 149* $12.50

Managerial Hierarchies

Comparative Perspectives on the Rise of the Modern Industrial Enterprise

Alfred D. Chandler, Jr., and Herman Daems, Editors

These essays show that the growth and complexity of managerial hierarchies in large businesses are central to the organization of modern industrial activity. This fresh study of the managerial revolution presents recent theoretical reflections in institutional economics and industrial organization in the light of new historical findings. *Harvard Studies in Business History, 32* $16.50

Now in paper—

The Nature of Mass Poverty

John Kenneth Galbraith

"A thoughtful, iconoclastic view of the subject: Galbraith's central point about the systemic nature of mass poverty makes it an excellent first word on the subject, and, as always, he presents his material with elegance and wit."—*Washington Post*
Paper, $2.95; cloth, $8.95

New Paperbacks—

Bankers and Pashas

International Finance and Economic Imperialism in Egypt

With a new Preface by the author

David S. Landes

"A valuable contribution to history, finance, diplomacy, and best of all, delightful reading."
—*Journal of Finance*

Paper, $6.95

Beyond Economic Man

A New Foundation for Microeconomics

With a new Preface by the author

Harvey Leibenstein

"In this major work, which every thinking economist should read,... Leibenstein has taken us a very long way down the road to a micro-theory that is more consistent with human behavior than the conventional theory."—*Journal of Economic Literature* Paper, $6.95; cloth, $15.00

HARVARD UNIVERSITY PRESS

Cambridge, Massachusetts 02138

The Efficient Use of Energy Resources

William D. Nordhaus

In what will undoubtedly be an important and influential book, a former member of the President's Council of Economic Advisers discusses the fundamental economic question relating to the use of energy resources: how fast should we consume our low-cost resources? With quantitative estimates developed through the use of a model, Nordhaus provides a set of forecasts and evaluations of the market-determined and economically efficient paths of energy production and consumption over time. $17.50

Money and Banking in Contemporary Japan

The Theoretical Setting and Its Application

Yoshio Suzuki

translated by John G. Greenwood with a foreword by James Tobin and Hugh Patrick

The translation into English of this prize-winning Japanese book is a major contribution to our understanding of Japanese financial institutions and monetary policy. Through the use of econometric estimation and other empirical analyses, Suzuki applies the theoretical framework developed by James Tobin and William Brainard to the actual circumstances of Japan and finds that the Tobin-Brainard model conforms well to the Japanese data; that is, Japan responds to income and price signals in much the same way other market-oriented economies do. $18.00

Arrow's Theorem: The Paradox of Social Choice

A Case Study in the Philosophy of Economics

Alfred F. MacKay

"A deep and very subtle analysis of a first-class problem. It is careful, sensitive, and just in dealing with its main target: Kenneth Arrow's celebrated theorem. This theorem has generated a brilliantly technical literature over a period of nearly three decades. But none of the previous commentaries really did anything to dispel the sense of mystery and surprise in the theorem. Here, for the first time, I find a way to really *understand* the mystery, and I am grateful to MacKay for having taught it to me."—Douglas W. Rae, Yale University

$14.50

Sweden: The Middle Way on Trial

Marquis W. Childs

Marquis Childs, a Pulitzer Prize-winning journalist and author of the now-classic *Sweden: The Middle Way,* takes another look at Sweden—forty years later—to see how well the "middle way" is faring. The Swedes, he finds, are grappling now with new problems, and the choices that must be made—about the costs of social welfare, about the distribution of wealth, and about developing nuclear power—are the choices that face the United States and other postindustrial states as well. Illus. $12.95

Yale University Press

New Haven and London

THE SOCIAL SYSTEM OF THE PLANET EARTH

by Kenneth B. Boulding, University of Colorado, Elise Boulding, Dartmouth College, and Guy M. Burgess, Massachusetts Institute of Technology

Innovative! Engaging! A unique interdisciplinary treatment of the world social system. A science fiction framework postulates the appearance of an extraterrestrial visitor who seeks knowledge of the Earth. Students get involved in organizing a research institute to gather and interpret data on population dynamics, history, religion, science, economics, social stratification, distribution of wealth, the international system, and more for this visitor.

AVAILABLE NOW (00207) paper 1980
Instructor's Manual *(00208)*

For complimentary copy consideration, write to Abigail Winslow. Be sure to mention course title, enrollment, and text now in use.

Social Science & Humanities Division
Addison-Wesley Publishing Company
Reading, Massachusetts 01867

PUBLICATIONS FROM THE UNITED NATIONS

WORLD INDUSTRY SINCE 1960: PROGRESS AND PROSPECTS

Special issue of the Industrial Development Survey for the Third General Conference of UNIDO (New Delhi, 21 Jan.–8 Feb. 1980)

Serves as one of the basic papers of the Conference; documents and analyzes with statistical tables the basic trends in the major areas of industry — production, trade, employment, technology and finance — and evaluates the implications of the changing relationship between the developing and developed countries.

Sales No. E.79.II.B.3 $18.00

WORLD ENERGY SUPPLIES 1973-1978

Summarizes world energy data on production, imports, exports and consumption and shows origin and destination of world trade in solid fuels, crude petroleum and natural gases.

Sales No. E.78.XVII.7 $14.00

UNITED NATIONS PUBLICATIONS

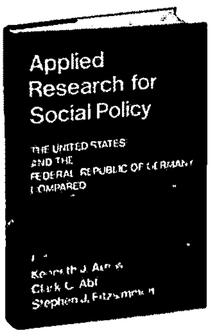
Room A-3315
New York, N.Y. 10017

Palais des Nations
1211 Geneva 10, Switzerland

Abt Books

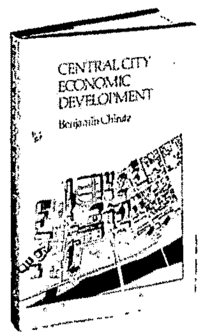
America's Newest Social Science Publisher

presents



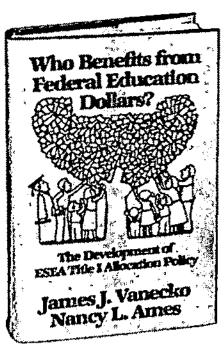
Kenneth J. Arrow,
Clark C. Abt,
Stephen J. Fitzsimmons,
editors,
Applied Research for Social Policy: The United States and the Federal Republic of Germany Compared, $30.00

Leading German and American social scientists compare notes on applied social research efforts and their impacts on decision makers in their respective countries.



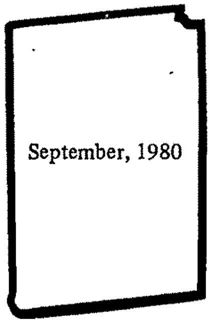
Benjamin Chinitz,
editor,
Central City Economic Development, $19.50

The nation's leading urban economists assess the overall effect of the 1970's on America's cities: job loss, erosion of tax base, inflation's impact. Government policy makers and heads of key economic and community development organizations contribute to the set of recommendations for aid to the cities.



James J. Vanecko
and Nancy L. Ames
with
Francis X. Archambault, Jr.,
Who Benefits from Federal Education Dollars? The Development of ESEA Title I Allocation Policy, $27.00

The policy question is how to distribute federal compensatory education funds most equitably; this book documents thirteen different experimental models over a three year period.



William S. Hendon,
James L. Shanahan,
Alice J. MacDonald, editors,
Economic Policy for the Arts

Experts from both sides of the Atlantic examine the economic aspects of the arts, issues like: private museums and public goods, the fine arts' contribution to commercial design, the performing arts and implications of government subsidies.

Abt Books

Dept. AER-5

55 Wheeler Street · Cambridge, Massachusetts · 02138

Tel.(617) 492-7100, ext. 337

Please mention The American Economic Review *When Writing to Advertisers*

EXECUTIVE SEARCH

SPECIALISTS IN:

ECONOMICS
ECONOMETRICS
CORPORATE PLANNING
STRATEGIC PLANNING
MARKETING
OPERATIONS RESEARCH

STONEHILL IS AN EXECUTIVE SEARCH FIRM STAFFED WITH A TEAM OF PROFESSIONALS, EXPERIENCED IN SEARCH AND RECRUITMENT TECHNIQUES.

WE TAKE PRIDE IN OUR NATIONWIDE REPUTATION FOR HIGH STANDARDS OF PROFESSIONALISM IN EXECUTIVE RECRUITING.

OUR CLIENTS ARE "FORTUNE 500" COMPANIES, MAJOR CORPORATIONS, FINANCIAL INSTITUTIONS, AND CONSULTING FIRMS.

CONTACT: MR. GARY J. STEVENS, PRESIDENT
MS. DEE DROST, V.P. EXECUTIVE SEARCH

STONEHILL
MANAGEMENT CONSULTANTS, INC.
521 FIFTH AVENUE
NEW YORK, NEW YORK 10017

212 687-8150 **SUITE 1760**

INTERNATIONAL ENERGY STRATEGIES

Proceedings of the 1979 IAEE/RFF Conference

Edited by Joy Dunkerley, *Resources for the Future*
(Published for the International Association of Energy Economists and Resources for the Future)
Prominent energy authorities discuss important questions (i.e. supply and demand, policy issues and choices, etc.) in the areas of energy demand and conservation, petroleum, other fuels (including coal and nuclear energy), and electricity.
Available - June 1980 ca. 384 pages ca. $25.00

INDUSTRY TIES WITH SCIENCE AND TECHNOLOGY POLICIES IN DEVELOPING COUNTRIES

Jack N. Behrman, *University of North Carolina*
(Published for the Fund for Multinational Management Education)
This study of science and technology (S&T) policies and organization in developing countries provides valuable background data for U.S. government and transnational corporation policies on S&T in less developed countries.
Available 224 pages $22.50

OVERSEAS R&D ACTIVITIES OF TRANSNATIONAL COMPANIES

Jack N. Behrman and William A. Fischer, *University of North Carolina*
(Published for the Fund for Multinational Management Education)
An indepth outline of various elements of the foreign R&D activities of transnational corporations and how these affect the host country.
Available ca. 336 pages $27.50

SCIENCE AND TECHNOLOGY FOR DEVELOPMENT

Corporate and Government Policies and Practices

Jack N. Behrman and William A. Fischer, *University of North Carolina*
(Published for the Fund for Multinational Management Education)
A survey of the issues to be considered in the formation of U.S. policies that will affect the development of science and technology capabilities in the less-developed countries.
Available 160 pages $20.00

THE TRANSFER OF TECHNOLOGY TO SOCIALIST COUNTRIES

The Case of the Soviet Chemical Industry

V. Sobeslavsky and P. Beazley
(Published for The Royal Institute of International Affairs, England, by Gower Publishing Co. Published in the United States by O, G & H)
A detailed examination of the use and structure of Western technology in expanding and modernizing the Russian chemical and petrochemical industries. The study emphasizes the role of political and economic relations in the transfer of technology between East and West.
176 pages ca. $21.00

HUMAN NEEDS: A CONTRIBUTION TO THE CURRENT DEBATE

Edited by Katrin Lederer, *Science Center Berlin*, in cooperation with Johan Galtung, *United Nations University*, and David Antal, *Science Center Berlin*
Internationally recognized experts evaluate current research on human needs. These papers, first presented at a conference at the International Institute for Environment and Society, Science Center Berlin, provide numerous insights into the theoretical, methodological, and practical implications of the concept of human needs.
Available ca. 320 pages ca. $25.00

INDICATOR SYSTEMS FOR POLITICAL, ECONOMIC, AND SOCIAL ANALYSIS

Edited by Charles Lewis Taylor, *Virginia Polytechnic Institute and State University*
These papers, the product of a conference at the International Institute for Comparative Social Research at the Science Center, Berlin, review the current state of indicator construction and systemization and examine the development and applications of quantitative indicators.
Available - June 1980 ca. 288 pages ca. $22.50

DECENTRALIZATION

Toward A Rational Theory

Manfred Kochen, *University of Michigan*, and Karl W. Deutsch, *Harvard University*
In this book, the authors describe a theory of systematic planning for the future of large service organizations incorporating general ideas about decentralization and stressing the identifiably common aspects of different service networks. They develop, organize, and analyze basic concepts and mathematical models which can explain the social and organizational effects of current trends toward more technology and demands for services.
Available — June 1980 ca. 400 pages ca. $27.50

NEW YORK — WORLD CITY

Report of the Twentieth Century Fund Task Force on the Future of New York City

Background Paper by Masha Sinnreich
Foreword by M. J. Rossant
This report assesses the resources of New York City and presents proposals to assure the future progress of the city and promote its status as a world capital. These proposals are designed to develop the resources of the city so that it can be an attractive and productive home for its citizens while emerging as the principal marketplace and cultural center of the world.
Available 256 pages cloth $17.50 paper $7.95

These titles are available from the publisher.

Telephone: (617) 876-5100

Oelgeschlager, Gunn & Hain, Inc.
Publishers
1278 Massachusetts Ave., Harvard Square, Cambridge, MA 02138
O G & H

New books from Resources from the Future

The Development of the U.S. Urban System

Concepts, Structure, Regional Shifts

Edgar S. Dunn, Jr.

This book is the first comprehensive quantitative history of the development of the U.S. urban system between the years 1940 and 1970. Written by Edgar S. Dunn, one of the pioneers in the field of urban economics, it challenges the popular theories of urban development and proposes the first step toward an alternative system.

Using a massive data base developed with the assistance of the Bureau of Economic Analysis, Dunn has matched the census data for employment by industry with such traditional descriptions of urban characteristics as commuting fields and functional trade center classifications. Eleven color plates, 25 full-page maps, 47 text tables, 7 figures, and the 421 pages of appendix tables on 10 microfiche cards can be used by readers to shape the data into regional configurations that are of special interest to them. $65.00

To Choose a Future

Resources and Environmental Consequences of Alternative Growth Paths

Ronald G. Ridker and William D. Watson

How will population growth, economic growth, and other major developments during the next fifty years affect the demand for resources? What environmental pressures will changes in these various factors cause?

The method employed here in answering these questions involves the preparation of scenarios—specially constructed sets of assumptions about future developments—and the study of their consequences. What emerges is a number of conditional projections that help sort out priorities and clarify the nature of the problems that must be faced. $33.50

plus one of the top energy books of the year

Energy in America's Future

The Choices Before Us

A Study Prepared for the RFF National Energy Strategies Project

Sam H. Schurr, Project Director

Joel Darmstadter, William Ramsay, Harry Perry, Milton Russell

"The best energy study yet to come along.... It is a veritable bible of information on energy, ticking off the advantages and disadvantages of the choices available."—*Washington Post* $30.00 hardcover; $10.95 paperback

Johns Hopkins

The Johns Hopkins University Press
Baltimore, Maryland 21218

OUTSTANDING Economics Texts from Addison-Wesley

In the forefront for 1980

Economics, Third Edition
by Lewis C. Solmon
A solid, systematic, successful new edition. Greatly reorganized, new design, clear writing style, superb coverage of theory and current real-world illustrations, policy chapters, applications, numerous built-in study aids. Extensive, coordinated teaching/learning package.
(07635), 1980, $17.95 (tent.)
(07217), **Macroeconomics,** *paper*
(07218), **Microeconomics,** *paper*

The New International Economics
by Jan S. Hogendorn, Colby College, and Wilson B. Brown, Northern Illinois University
Stimulating new text explains the historical, political, and institutional background and emphasizes the importance of a variety of major topics. Continually links theory with policy issues. Includes Swan-Mundell diagrams, Mead tables, and discusses new theories of trade by Linder, Wells-Vernon, and Grubel.
(02824), 1979, $17.25

Managerial Economics
by Pearson Graham, TRW, Inc. and Cleveland State University, and Diran Bodenhorn, Cleveland State University
(02470), 1980, $16.95

The Social System of the Planet Earth
by Kenneth E. Boulding, University of Colorado, Elise Boulding, Dartmouth College, and Guy M. Burgess, Massachusetts Institute of Technology
(00207), 1980, paper, $9.95 (tent.)

Some of the Best on Backlist

Economics: An Introduction to the World Around You, Second Edition
by Dennis J. Weidenaar and Emanuel T. Weiler, both of Purdue University
(08517), 1979, paper, $13.95

Economics: A Tool for Understanding Society
by Tom Riddell, Steve Stamos, and Jean Shackelford, all of Bucknell University
(06352), 1979, paper, $11.95

Economics of Exchange Rates: Selected Studies
by Jacob A. Frenkel, University of Chicago, and the late Harry G. Johnson, formerly University of Chicago
(02374), 1978, $13.95
(02376), paper, $9.50

Capitalism's Inflation and Unemployment Crisis: Beyond Monetarism and Keynesianism
by Sidney Weintraub, University of Pennsylvania
(08503), 1978, $10.75
(08502), paper, $7.50

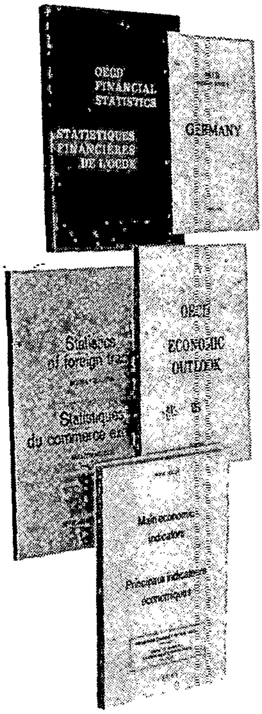
For complimentary copy consideration, write to Abigail Winslow. Be sure to mention course title, enrollment, and text now in use.

Social Science & Humanities Division
Addison-Wesley Publishing Company
Reading, Massachusetts 01867

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

OECD

OECD is the world's largest group of industrialized market-economy countries, comprising nineteen European nations, the U.S.A., Canada, Japan, Australia and New Zealand. The Organization publishes a unique and timely range of comparative international data for businessmen, government officials and the academic world.



OECD ECONOMIC OUTLOOK

Each July and December the OUTLOOK surveys the latest economic developments in the OECD area and, by means of an integrated set of quantitative forecasts, assesses future prospects. Reserve your July copy now, $8.75.
Annual subscription **$17.50** By air **$19.90**

MAIN ECONOMIC INDICATORS

Provides at a glance a picture of the most recent changes in the economy of the OECD countries. The indicators selected cover national accounts, industrial production, deliveries, stocks and orders, construction, retail sales, labor, wages, prices, domestic and foreign finance, interest rates, trade and payments. Monthly issue **$5.00** Subscription **$50.00** By air **$71.30**

OECD ECONOMIC SURVEYS

Each country survey provides a detailed analysis of recent developments in demand, production, employment, prices and wages, conditions in the money and capital markets, and developments in the balance of payments. They include short-term forecasts as well as analyses of medium-term problems. **$3.50** per country. Annual series subscription **$60.00** By air **$76.90**

OECD FINANCIAL STATISTICS

A unique collection of statistical and descriptive data on the international financial market and the national financial markets of sixteen European countries, the United States, Canada and Japan. Supplements give between the 15th and 20th each month, more than 300 interest rates recorded the previous month on the national markets and the international market.
Annual subscription **$95.00** By air **$118.30**

STATISTICS OF FOREIGN TRADE Series A. *Monthly Bulletin*

Shows total trade statistics for each country by partner country, as well as totals for main regions. It also gives indices of volume and average value by country, together with a breakdown of foreign trade values by categories based on the SITC. Subscription **$45.00** By air **$64.20**

FACING THE FUTURE: MASTERING THE PROBABLE AND MANAGING THE UNPREDICTABLE (Interfutures Report) The Interfutures Project, launched over three years ago, has attempted to get an independent assessment of the problems which the advanced industrial societies will face in the long term, both in their own development and in their relations with developing countries.
(September 1979) 425 pp. **$20.00**

REVENUE STATISTICS OF OECD MEMBER COUNTRIES 1965-1978 Detailed tables, by country, on sources of government revenues, including compulsory social security contributions. Special annex on the allocation of tax revenues by levels of government. Updated yearly. (September 1979) 235 pp. **$17.00**

EQUAL OPPORTUNITIES FOR WOMEN What are the consequences of the influx of women on the labor market? The study focuses on six themes, illustrated by specific country experiences. which should encourage innovation elsewhere.
(April 1979) 216 pp. **$10.00**

THE IMPACT OF THE NEWLY INDUSTRIALIZING COUNTRIES ON PRODUCTION AND TRADE IN MANUFACTURES Using an illustrative sample of 10 middle-income developing countries, this report discusses the problems and the benefits of increased trade in both directions.
(June 1979) 96 pp. **$12.00**

INTERNATIONAL INVESTMENT AND MULTINATIONAL ENTERPRISES.
Review of the 1976 Declaration and Decisions
Provides explanatory comments on the guidelines in the light of experience. (August 1979) 66 pp. **$6.00**

INTERNATIONAL INVESTMENT AND MULTINATIONAL ENTERPRISES: REVISED 1979 EDITION The new version of the OECD "Guidelines," which resulted from the June 1979 review mentioned above. (December 1979) 24 pp. **$3.00**

*Titles may be ordered individually. For faster delivery at lower cost you can now place a **Selective Standing Order** in your area of interest. All OECD publications are also available on **microfiche.** Write for free Catalog:*

OECD PUBLICATIONS AND INFORMATION CENTER
1750-E PENNSYLVANIA AVE., N.W. WASHINGTON, D.C. 20006 (202) 724-1857
Also available through U.N. Bookshop, New York and sales agents throughout the world.

AMERICAN ECONOMIC ASSOCIATION

Organized at Saratoga, New York, September 9, 1885

PAST OFFICERS

Presidents

FRANCIS A. WALKER,* M.I.T., 1886–92
CHARLES F. DUNBAR,* Harvard, 1893
JOHN B. CLARK,* Columbia, 1894–95
HENRY C. ADAMS,* Michigan, 1896–97
ARTHUR T. HADLEY,* Yale, 1898–99
RICHARD T. ELY,* Wisconsin, 1900–01
EDWIN R. A. SELIGMAN,* Columbia, 1901–03
FRANK W. TAUSSIG,* Harvard, 1904–05
JEREMIAH W. JENKS,* Cornell, 1906–07
SIMON N. PATTEN,* Pennsylvania, 1908
DAVIS R. DEWEY,* M.I.T., 1909
EDMUND J. JAMES,* Illinois, 1910
HENRY W. FARNUM,* Yale, 1911
FRANK A. FETTER,* Princeton, 1912
DAVID KINLEY,* Illinois, 1913
JOHN H. GRAY,* Minnesota, 1914
WALTER F. WILLCOX,* Cornell, 1915
THOMAS N. CARVER,* Harvard, 1916
JOHN R. COMMONS,* Wisconsin, 1917
IRVING FISHER,* Yale, 1918
HENRY B. GARDNER,* Brown, 1919
HERBERT J. DAVENPORT,* Cornell, 1920
JACOB H. HOLLANDER,* Johns Hopkins, 1921
HENRY R. SEAGER,* Columbia, 1922
CARL C. PLEHN,* California, 1923
WESLEY C. MITCHELL,* Columbia, 1924
ALLYN A. YOUNG,* Harvard, 1925
EDWIN W. KEMMERER,* Princeton, 1926
THOMAS S. ADAMS,* Yale, 1927
FRED M. TAYLOR,* Michigan, 1928
EDWIN F. GAY,* Harvard, 1929
MATTHEW B. HAMMOND,* Ohio State, 1930
ERNEST L. BOGART,* Illinois, 1931
GEORGE E. BARNETT,* Johns Hopkins, 1932
WILLIAM Z. RIPLEY,* Harvard, 1933
HARRY A. MILLIS,* Chicago, 1934
JOHN M. CLARK,* Columbia, 1935
ALVIN S. JOHNSON,* New School, 1936
OLIVER M. W. SPRAGUE,* Harvard, 1937
ALVIN H. HANSEN,* Harvard, 1938
JACOB VINER,* Chicago, 1939
FREDERICK C. MILLS,* Columbia, 1940
SUMNER H. SLICHTER,* Harvard, 1941
EDWIN G. NOURSE,* Brookings, 1942
ALBERT B. WOLFE,* Ohio State, 1943
JOSEPH S. DAVIS,* Stanford, 1944
I. L. SHARFMAN,* Michigan, 1945
E. A. GOLDENWEISER,* Inst. Advanced Study, 1946
PAUL H. DOUGLAS,* Chicago, 1947
JOSEPH A. SCHUMPETER,* Harvard, 1948
HOWARD S. ELLIS, California, 1949
FRANK H. KNIGHT,* Chicago, 1950
JOHN H. WILLIAMS, Harvard, 1951
HAROLD A. INNIS,* Toronto, 1952
CALVIN B. HOOVER,* Duke, 1953
SIMON KUZNETS, Pennsylvania, 1954
JOHN D. BLACK,* Harvard, 1955
EDWIN E. WITTE,* Wisconsin, 1956
MORRIS A. COPELAND, Cornell, 1957
GEORGE W. STOCKING,* Vanderbilt, 1958
ARTHUR F. BURNS, Columbia, 1959
THEODORE W. SCHULTZ, Chicago, 1960
PAUL A. SAMUELSON, M.I.T., 1961
EDWARD S. MASON, Harvard, 1962
GOTTFRIED HABERLER, Harvard, 1963
GEORGE J. STIGLER, Chicago, 1964
JOSEPH J. SPENGLER, Duke, 1965
FRITZ MACHLUP, Princeton, 1966
MILTON FRIEDMAN, Chicago, 1967
KENNETH E. BOULDING, Colorado, 1968
WILLIAM J. FELLNER, Yale, 1969
WASSILY LEONTIEF, Harvard, 1970
JAMES TOBIN, Yale, 1971
JOHN KENNETH GALBRAITH, Harvard, 1972
KENNETH J. ARROW, Harvard, 1973
WALTER W. HELLER, Minnesota, 1974
ROBERT AARON GORDON,* California, 1975
FRANCO MODIGLIANI, M.I.T., 1976
LAWRENCE R. KLEIN, Pennsylvania, 1977
JACOB MARSCHAK,* California, 1978†
TJALLING C. KOOPMANS, Yale, 1978
ROBERT M. SOLOW, M.I.T., 1979

Secretaries

RICHARD T. ELY,* 1886–92
EDWARD A. ROSS,* 1893
JEREMIAH W. JENKS,* 1894–96
WALTER F. WILLCOX,* 1897–99

Treasurers

EDWIN R. A. SELIGMAN,* 1886–90
FREDERICK B. HAWLEY,* 1891–95
CHARLES H. HULL,* 1896–99

Secretary-Treasurers

CHARLES H. HULL,* 1900
FRANK A. FETTER,* 1901–06
WINTHROP M. DANIELS,* 1907–08
THOMAS N. CARVER,* 1909–13
ALLYN A. YOUNG,* 1914–20
RAY B. WESTERFIELD,* 1921–25
FREDERICK S. DEIBLER,* 1925–35
JAMES WASHINGTON BELL,* 1936–61
HAROLD F. WILLIAMSON, 1962–70
RENDIGS FELS, 1971–75

Editors

DAVIS R. DEWEY,* 1911–40
PAUL T. HOMAN,* 1941–51
BERNARD F. HALEY, 1952–62
JOHN G. GURLEY, 1962–68
ARTHUR SMITHIES, 1963–68

*Deceased
†Died before taking office

Economic Update

Just Published

PRINCIPLES OF ECONOMICS: MACRO, Fourth Edition
PRINCIPLES OF ECONOMICS: MICRO, Fourth Edition
Willis L. Peterson, University of Minnesota
These **twin paperbound volumes** are linked to each other, although they are designed to be used in either a macro-micro or reverse sequence. A **new Study Guide** is available for each volume.

THE PUBLIC FINANCES: An Introductory Textbook, Fifth Edition
James M. Buchanan, Virginia Polytechnic Institute and State University, and **Marilyn R. Flowers,** University of Oklahoma

MONEY, BANKING, AND MONETARY POLICY
Douglas Fisher, North Carolina State University

PROBLEMS IN LABOR RELATIONS: Text and Cases
William E. Fulmer, University of Alabama

MICROECONOMIC THEORY, Fifth Edition
John P. Gould, University of Chicago, and the late **C. E. Ferguson**

THE ECONOMICS OF DEVELOPMENT, Third Edition
Everett E. Hagen, Emeritus, Massachusetts Institute of Technology

LABOR ECONOMICS: Wages, Employment and Trade Unionism, Fourth Edition
F. Ray Marshall, U.S. Secretary of Labor, **Allan G. King,** The University of Texas at Austin, and **Vernon M. Briggs, Jr.,** Cornell University

READINGS IN LABOR ECONOMICS AND LABOR RELATIONS: Fourth Edition *(paperbound)*
Edited by **Richard L. Rowan,** University of Pennsylvania

Examination copies for adoption consideration available on request; please indicate course title and text presently used.

RICHARD D. IRWIN, INC. • Homewood, Illinois 60430